PLOS GENETICS

RESEARCH ARTICLE

Comparing DNA replication programs reveals large timing shifts at centromeres of endocycling cells in maize roots

Emily E. Wear[1]*, Jawon Song[2], Gregory J. Zynda[2], Leigh Mickelson-Young[1], Chantal LeBlanc[3¤a], Tae-Jin Lee[1¤b], David O. Deppong[1], George C. Allen[4], Robert A. Martienssen[3], Matthew W. Vaughn[2], Linda Hanley-Bowdoin[1], William F. Thompson[1]

1 Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, United States of America, **2** Texas Advanced Computing Center, University of Texas, Austin, Texas, United States of America, **3** Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America, **4** Department of Horticultural Science, North Carolina State University, Raleigh, North Carolina, United States of America

¤a Current address: Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut, United States of America
¤b Current address: Syngenta Crop Protection, Research Triangle Park, North Carolina, United States of America
* emily_wear@ncsu.edu

Abstract

Plant cells undergo two types of cell cycles–the mitotic cycle in which DNA replication is coupled to mitosis, and the endocycle in which DNA replication occurs in the absence of cell division. To investigate DNA replication programs in these two types of cell cycles, we pulse labeled intact root tips of maize (*Zea mays*) with 5-ethynyl-2'-deoxyuridine (EdU) and used flow sorting of nuclei to examine DNA replication timing (RT) during the transition from a mitotic cycle to an endocycle. Comparison of the sequence-based RT profiles showed that most regions of the maize genome replicate at the same time during S phase in mitotic and endocycling cells, despite the need to replicate twice as much DNA in the endocycle and the fact that endocycling is typically associated with cell differentiation. However, regions collectively corresponding to 2% of the genome displayed significant changes in timing between the two types of cell cycles. The majority of these regions are small with a median size of 135 kb, shift to a later RT in the endocycle, and are enriched for genes expressed in the root tip. We found larger regions that shifted RT in centromeres of seven of the ten maize chromosomes. These regions covered the majority of the previously defined functional centromere, which ranged between 1 and 2 Mb in size in the reference genome. They replicate mainly during mid S phase in mitotic cells but primarily in late S phase of the endocycle. In contrast, the immediately adjacent pericentromere sequences are primarily late replicating in both cell cycles. Analysis of CENH3 enrichment levels in 8C vs 2C nuclei suggested that there is only a partial replacement of CENH3 nucleosomes after endocycle replication is complete. The shift to later replication of centromeres and possible reduction in CENH3 enrichment after endocycle replication is consistent with a hypothesis that centromeres are inactivated when their function is no longer needed.

Author summary

In traditional cell division, or mitosis, a cell's genetic material is duplicated and then split between two daughter cells. In contrast, in some specialized cell types, the DNA is duplicated a second time without an intervening division step, resulting in cells that carry twice as much DNA. This phenomenon, which is called the endocycle, is common during plant development. At each step, DNA replication follows an ordered program in which highly compacted DNA is unraveled and replicated in sections at different times during the synthesis (S) phase. In plants, it is unclear whether traditional and endocycle programs are the same, especially since endocycling cells are typically in the process of differentiation. Using root tips of maize, we found that in comparison to replication in the mitotic cell cycle, there is a small portion of the genome whose replication in the endocycle is shifted in time, usually to later in S phase. Some of these regions are scattered around the genome and mostly coincide with active genes. However, the most prominent shifts occur in centromeres. The shift to later replication in centromeres is noteworthy because they orchestrate the process of separating duplicated chromosomes into daughter cells, a function that is not needed in the endocycle.

Introduction

Developmentally programmed DNA replication without nuclear breakdown, chromosome condensation or cell division, a phenomenon known as endoreduplication or endocycling, occurs in a wide variety of plants and animals [1–3]. In plants, endoreduplication is a systemic feature [4] that is often an important step in the development of tissues and organs such as fruit, endosperm, leaf epidermal cells and trichomes [5]. Initiation of endocycling is frequently associated with a transition from cell proliferation to cell differentiation and expansion [6]. In plant roots, cells at the tip divide actively by normal mitosis, while endocycling cells become frequent further from the tip, in a zone associated with differentiation and increased cell size [7, 8]. In the maize (*Zea mays*) root, the first 1-mm primarily contains actively dividing mitotic cells [8, 9]. Mitotic activity continues in the region between 1 to 3-mm from the tip, but in this region about 30% of the cells undergo a single endocycle. Instead of undergoing mitosis and returning to a 2C nuclear DNA content, these cells replicate their DNA without undergoing mitosis and transition directly from 4C to a final DNA content of 8C [8]. Beyond 3 mm, cell differentiation predominates, and DNA replication activity becomes rare [8, 10]. Thus, while not all differentiating cells undergo an endocycle, in those cells that do, the endocycle precedes or accompanies cellular differentiation.

In animal systems, different cell types often exhibit differences in the temporal order of DNA replication along the chromosomes, also called the "replication timing (RT) program". These differences in RT programs are likely related to changes in chromatin and gene expression during the differentiation process (e.g. [11–14]). However, as yet there is no information concerning possible changes in RT programs associated with endoreduplication or differentiation in plant systems. Other than in a few model systems (e.g. [15, 16]), it is difficult to separate individual cell types from plant organs, and DNA replication occurs mainly in meristematic regions, which are small and often difficult to access. In addition, plant cells do not maintain their differentiated state when grown in suspension culture [17, 18]. Hence, there are few opportunities to compare RT programs in individual cell types. However, the occurrence of endocycling and differentiating cells near the apical meristem of maize root tips, in which we

previously characterized the mitotic RT program [19], offers a unique opportunity to compare the two modes of replication in an intact plant organ.

We developed a system to analyze DNA replication in maize roots [8, 20], using similar approaches to those being applied in our work with *Arabidopsis* cell suspensions [21]. In this system, newly replicated DNA is labeled *in vivo* with the thymidine analog, 5-ethynyl-2'-deoxy-yuridine (EdU), and labeled nuclei are separated by flow cytometry into populations representing different stages of S phase. One key advantage of using EdU as the nucleotide analog instead of 5-bromo-2'deoxyuridine (BrdU), which has traditionally been used in replication timing by sequencing (Repli-seq) protocols [12], is the ability to sort nuclei on both DNA content as well as EdU content, followed by immunoprecipitation of EdU labeled DNA. This modification allows for clean separation during flow sorting of non-replicating, unlabeled G1 and G2 nuclei from the labeled S-phase nuclei.

Cytological analysis of sorted, EdU-labeled nuclei showed that replication activities in early and mid S are more closely interspersed in the maize nucleus than in animal cells [22, 23]. We characterized the RT program in mitotic cells of the apical 1-mm root segment [19], using the modified Repli-seq protocol [24]. In mitotic cells, we found evidence for a gradient of early replicating, open chromatin that transitions gradually into less open and less transcriptionally active chromatin replicating in mid S phase. We also confirmed cytological observations showing that heavily compacted classical heterochromatin, including knobs and pericentromeres, replicate primarily in late S phase [22, 25]. While these relationships between RT and chromatin packaging are generally similar to those found in other systems, we did not find evidence for megabase-scale replication domains like those that have been characterized in mammalian cells (reviewed in [26] and references therein).

Although replication in the first 1-mm of the root is mostly mitotic, with DNA contents of labeled nuclei ranging from 2C to 4C, flow cytometry profiles of nuclei derived from root tissue between 1 and 3-mm from the tip also included a substantial population of nuclei with DNA contents beyond 4C. Appearance of endopolyploid (8C) nuclei in this zone would be expected, as some 4C nuclei are known to enter the endocycle rather than undergo division [7, 8]. Cytological analysis of replicative labeling showed that the spatiotemporal patterns of replication during the 4C to 8C transition in these endocycling nuclei are very similar to those in mitotic nuclei [22]. However, it remained to be determined whether the entire genome is uniformly replicated during the endocycle, and whether the temporal program is altered in differentiating cells when replication occurs without an intervening mitosis.

Both under-replication and over-replication (amplification) have been observed in multiple animal systems during developmentally programmed endocycles, notably including *Drosophila* (reviewed in [27]). In addition to the well-known amplification of chorion genes and under-replication of heterochromatin, under-replication also occurs in a number of euchromatic regions, with a degree of tissue specificity suggesting a possible role in differentiation [28–30].

Even though endopolyploidy is common in plants, there are very few reports dealing with over- or under-replication of specific sequences. Some orchids exhibit a phenomenon in which only a fraction of the genome is endoreplicated [31, 32], but in most cases, endopolyploid cells have DNA contents that are multiples of the 2C value. Both highly repetitive heterochromatic regions and highly expressed genes are extensively endoreduplicated in maize endosperm nuclei, as would be expected for uniform replication of the entire genome [33]. More definitively, whole genome sequencing in *Arabidopsis* showed that leaf nuclear DNA is evenly endoreduplicated in wild-type plants, although the same series of experiments clearly demonstrated selective over-replication in *atxr5* and *atxr6* mutants [34].

To address the question of over- or under-replication and whether there are developmentally associated changes in RT programs in the maize root tip system, we performed a detailed comparison of RT dynamics in mitotic and endocycling cells. We found very little evidence for over- or under-replication in endocycling cells, consistent with the few previous reports on this topic from plant systems. We also found that the RT programs for the vast majority of the genome are very similar. However, we found significant changes in timing for a number of loci that together correspond to 2% of the genome. Most notably, we found major changes in the RT of centromeres, which replicate mainly during mid S phase in mitotic cells but primarily in late S phase of the endocycle.

Results

Separating endocycling from mitotic nuclei

As reported previously and described in Methods, we used a 20-min pulse of the thymidine analog, EdU, to label newly replicated DNA in intact maize roots. This was followed by formaldehyde fixation and isolation of nuclei from defined segments of root tips (Fig 1A). Incorporated EdU was conjugated with Alexa Fluor 488 (AF-488) by "click" chemistry [35]. The nuclei were then stained with DAPI and fractionated by two-color fluorescence activated flow sorting to generate populations at different stages of the mitotic cell cycle or the endocycle [8, 20]. Fig 1B and 1C show flow cytometry profiles obtained for root segments 0–1 mm and 1–3 mm from the tip, respectively. Fluorescent signals from nuclei that incorporated EdU during S phase of a normal mitosis form an "arc" between 2C and 4C DNA contents (Fig 1B). Previous EdU pulse-chase time course experiments in the 0–1 mm region showed that most labeled 4C nuclei return to 2C as would be expected in a mitotic cycle (see Fig 1B in [36]). In the 1–3 mm zone, a substantial fraction of labeled nuclei is undergoing endoreplication, forming a similar arc with DNA contents ranging between 4C and 8C (Fig 1C). As seen in Fig 1C, the endocycle arc is more prominent in nuclei preparations from 1–3 mm root segments. To analyze the endocycle RT program, which is described in detail below, labeled nuclei representing early, mid, and late S-phase fractions were separated using the sorting gates shown in Fig 1C, adjusting the endocycle early gate to avoid contamination with mitotic nuclei in late S phase. Reanalysis of the sorted nuclei confirmed that there was good separation between the nuclei populations from the adjusted early sorting gate and the mid sorting gate (S1 Fig). To complete the Repli-seq protocol, described in more detail below, the DNA was extracted from the nuclei in each gate and sheared. The labeled, newly replicated DNA from each S-phase fraction was then immunoprecipitated and sequenced.

The flexibility of the EdU labeling and flow sorting system also allowed us to collect unlabeled nuclei, representing non S-phase cells with 2C, 4C and 8C DNA contents. These nuclei were used to characterize selected histone marks following mitotic or endocycle replication and to investigate the copy number of individual loci across the genome.

Evidence for complete genome replication during the endocycle

Given the well documented examples of over- and under-replication during the endocycle in animal systems, we investigated whether there are local copy number differences in the maize genome after endocycle replication. To do this, we used the non S-phase 2C, 4C, and 8C nuclei populations described above, and carried out whole genome paired-end sequencing. To gain a better representation of the copy number of repeat regions in the genome, reads that could not be uniquely mapped to a single location were included, but we retained only the primary alignment location for each read pair. This approach was used exclusively for the copy number analysis, while all subsequent analyses included only uniquely mapping reads (see Methods
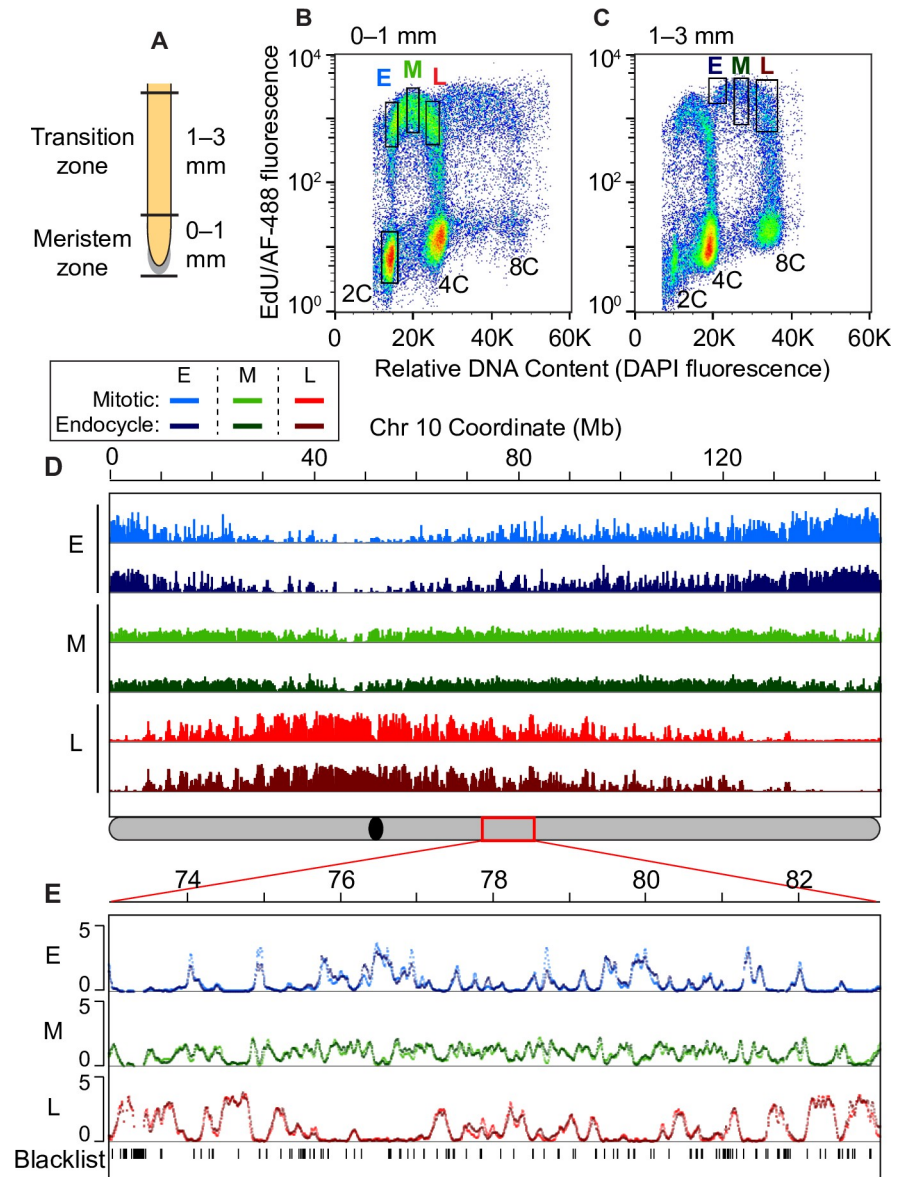
Fig 1. Global comparison of mitotic cycle and endocycle replication timing programs. (**A**) Schematic of a maize root showing the meristem zone (0–1 mm region) and transition zone (1–3 mm region) used for Repli-seq experiments. (**B** and **C**) Flow cytograms of nuclei isolated from the 0–1 mm root segments (**B**) and 1–3 mm root segments (**C**). Dots are pseudo-colored by density and black rectangles represent the sorting gates used to collect the pre-replicative 2C reference sample and early (E), mid (M) and late (L) S-phase fractions from either the mitotic cycle or endocycle. (**D**) Global scale view of replication timing (RT) for chromosome 10, comparing mitotic and endocycling profiles in early, mid and late S phase. Uniquely mapping reads were aggregated in 3-kb windows, normalized for sequencing depth, divided by the normalized 2C reference read counts, and Haar wavelet smoothed (see Methods). The global RT profiles for mitotic and endocycling cells are very similar to each other for all ten chromosomes. The schematic of chromosome 10 at the bottom shows the location of the centromere (black oval) and the 10 Mb region that is expanded in panel **E** (red rectangle). (**E**) Expanded view of a 10 Mb region on chromosome 10 with overlaid mitotic and endocycle RT profiles. Unmappable or multi-mapping regions ("blacklist") were identified from the pre-replicative 2C reference sample and are indicated as tick marks in the bottom track. This example illustrates the similarity between the mitotic and endocycle RT profiles that is observed throughout most of the genome. Scale for all panels: 0–5 normalized replication signal.

https://doi.org/10.1371/journal.pgen.1008623.g001

and S1 Text). The data for each ploidy level were examined for regions in which normalized read frequencies in 5-kb windows differed between 8C and 4C or 4C and 2C nuclei, using procedures described by Yarosh et al. ([37]; S1 Text). We found about 5% of the 5-kb windows had ratio values that fell outside of two standard deviations of the mean ratio for 4C and 2C or 8C and 4C (1.0 ± 0.2 S. D. for both; S2A and S2B Fig). However, these windows all either occurred as singleton 5-kb windows scattered around the genome (S2C Fig) or coincided with regions that had very low read mapping in the 2C sample, indicating they are likely the spurious result of making a ratio between windows with very few reads in both samples. As such, there is very little evidence of meaningful over- or under-replication of genomic regions in nuclei with different ploidy levels.

To further investigate whether there is complete replication of high-copy repeats that are not well represented in the genome assembly, we used BLAST software to query all reads, not just those that can be mapped to the genome, to determine the percentage of reads corresponding to each of several consensus sequences for high-copy repeats (S1 Text). Analyzed sequences included the knob repeats *knob180* and *TR-1* [38, 39], 5S and 45S rDNA repeats [40], and centromere-associated *CentC* satellite repeats [41]. We also queried consensus sequences for centromere retrotransposons of maize (*CRM*) families 1–4 [42–45]. In all cases, we found the percentages to be similar in the 2C, 4C and 8C samples (S2D and S2E Fig), further suggesting that there is little or no over- or under-replication.

Replication timing analysis

As described above, we sorted endocycling nuclei from the S-phase populations in Fig 1C, and extracted and sheared the DNA in each fraction. EdU-containing DNA fragments were immunoprecipitated (IP) with an antibody to AF-488, resulting in sequence populations representing DNA replicating during early, middle, or late S phase of the endocycle. Given that EdU/AF-488-labeled DNA is immunoprecipitated from a population of only EdU-labeled nuclei (see Fig 1B and 1C), the level of unlabeled DNA background in each IP is substantially lower than would be present if nuclei were sorted simply by DNA content. We also prepared DNA from the unlabeled 2C nuclei pool to provide a reference dataset representing pre-replicative nuclei. DNA from three biological replicates of each sample was sequenced to generate paired-end reads.

To compare the RT programs in endocycling and mitotic nuclei, we mapped our previous Repli-seq data for mitotic nuclei [19] and our new data for endocycling nuclei to the maize B73 RefGen_v4 genome, which includes improved assemblies of centromeres and more complete annotations of transposable elements (TEs) [46, 47]. Including only uniquely mapped reads resulted in a read depth that varied between 65.7 million and 261.2 million reads per S-phase fraction (including reads from 3 biological replicates per S-phase fraction). The data from all S-phase samples for the mitotic cycle and endocycle were then randomly downsampled to the lowest read depth (65.7 million reads) to ensure comparable results (see Methods and S1 Spreadsheet).

We used the *Repliscan* analysis pipeline [24] to generate continuous normalized data profiles representing the intensity of replication activity across the genome (RT profiles) in early, mid and late fractions of each S phase. These RT profiles were generated by aggregating the Repli-seq read densities for each S-phase sample in 3-kb static windows, scaling the reads to 1× genome coverage, and then dividing by the scaled read counts from the unlabeled 2C reference data and smoothing by Haar wavelet transform (see Methods and [24]). Normalizing with the pre-replicative 2C reference provided a uniform 2C copy number and corrected for differences in sequence mappability and collapsed repeats that caused "spikes" in the data

(illustrated for late replication in the endocycle in S3 Fig) while preserving replication signal [24, 48]. The 3-kb windows identified as having no or extremely low read coverage in the 2C reference sample (see Methods) were excluded from all analyses. These windows include both unmappable and multi-mapping regions ("blacklist" windows, indicated by black tick marks in Fig 1E). After these RT profile normalization steps, the result is an estimate of the intensity of replication activity in each 3-kb window, which we refer to as "replication signal".

Fig 1D shows that the global RT profiles are remarkably similar in endocycling and mitotic nuclei, and overlays of the corresponding profiles show mostly minor differences (Fig 1E). Pearson's correlation coefficient values between corresponding S-phase fractions from the mitotic and endocycle data are very high (r values of 0.91, 0.89 and 0.96 for early, mid and late, respectively). These values are similar to those found between individual biological replicates within each sample (S4 Fig).

Identifying regions of altered timing

Despite the global similarity of the RT programs of mitotic and endocycling cells, there are regions scattered around the maize genome that show a shift in RT. To identify regions with differences in RT (DRT), we first calculated the difference in normalized replication signal between the mitotic and endocycle data at each genomic location for the early, mid and late profiles separately (S1 Table; S5 Fig). We then constrained our analysis by focusing only on regions where there was an equal and opposite DRT in at least one other S-phase fraction (for example, regions in which a decrease in early replication signal in endocycling cells was associated with a corresponding increase in mid and/or late replication signal at the same location). We allowed a gap distance of 6 kb when searching for regions with DRT to account for small blacklist regions that break up larger regions of change. We found that 11% of the genome showed a difference in replication signal of at least 10% of the total difference range for a given S-phase fraction (absolute difference in replication signal ≥ 0.4; S1 Table), with an opposite difference in replication signal at the same threshold criterion at the identical location in another S-phase fraction. Many of these regions are small, with the lower 50% of regions ranging in size from 3 kb to the median size of 33 kb (S2 Table). Since the units of replication initiation, elongation and termination ("replicons") in monocot plants are estimated to be on average 47 ± 13 kb in size [49], the biological relevance of RT differences across regions much smaller than this is not clear. However, instead of implementing an arbitrary size cutoff, we chose to focus on regions that were associated with at least one core region with a larger (≥ 25%) DRT, as described in the next paragraph.

To identify these more robust RT differences, designated Regions of Altered Timing (RATs), we identified regions in which the DRT was ≥ 25% of the total difference range for a given S-phase fraction (absolute difference in replication signal ≥ 1.0; S1 Table), and which also met the criterion of having an opposite DRT in at least one other S-phase fraction. To highlight larger and contiguous regions of change, we included ≥ 10% regions that were adjacent to the original ≥ 25% regions. However, RATs had to have at least one core region where the DRT was at least 25% (S2 Table) to be included in our analysis. Representative ≥ 25% and ≥ 10% regions are indicated by various shades of red and blue bars in Fig 2 (additional examples are in S6 Fig). Finally, we examined the RT profiles for the RATs in individual biological replicates to verify there was good agreement between the replicates (Figs 2B and S6). By selecting only the most robust RATs we excluded other regions where RT changes are less dramatic–for example those indicated by dashed boxes in Fig 2. In such regions, the DRT did not meet our criteria of a ≥ 25% difference in replication signal (box 2 in Fig 2A) and/or there is not an equal and opposite ("compensated") DRT (box 3 in Fig 2A).
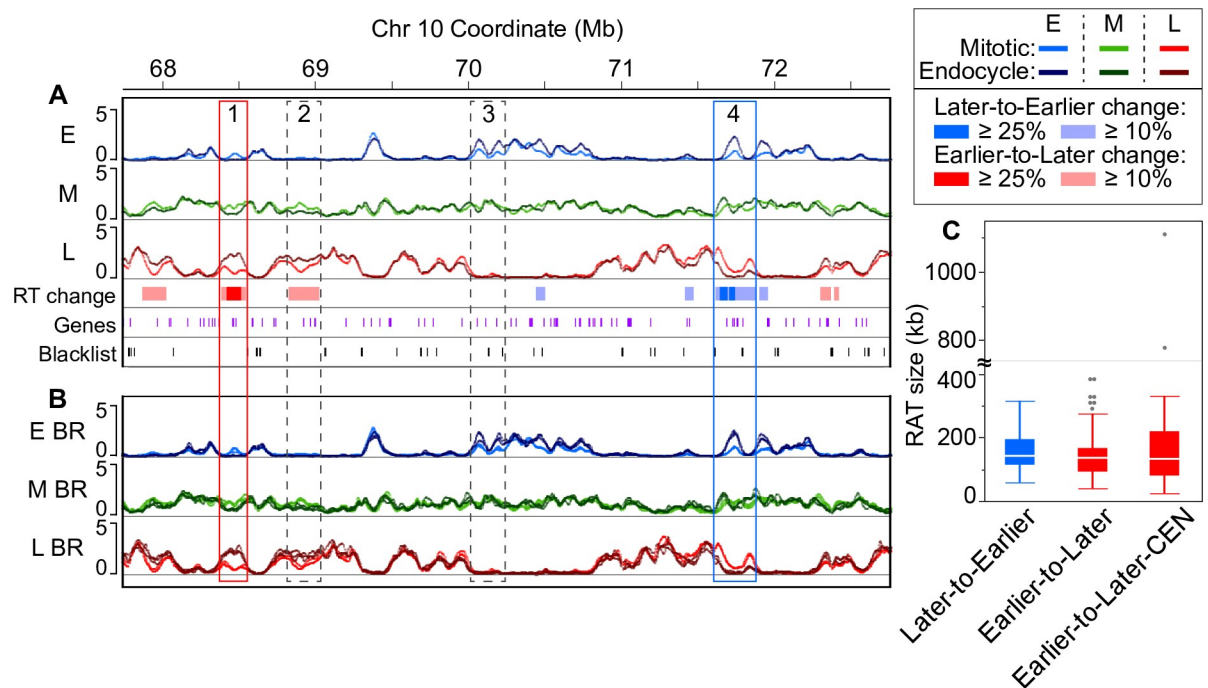
Fig 2. Identifying regions of altered timing. (**A**) An example region (5 Mb) on chromosome 10 containing two robust Regions of Altered Timing (RATs), indicated by boxes outlined with solid lines. The RAT in box 1 (red) shifts from Earlier-to-Later, and the RAT in box 4 (blue) shifts from Later-to-Earlier. Dashed boxes denote regions with some level of difference in RT (DRT) in which the magnitude of the difference did not meet our $\geq$ 25% criterion (box 2), or in which the change in one S-phase fraction was not compensated by an opposite change in at least one other S-phase fraction (box 3). Annotated genes (purple) and unmappable or multi-mapping regions ("blacklist", black) are indicated as tick marks in the bottom tracks. (**B**) The same chromosome region as in (**A**) with the individual biological replicate RT profiles overlaid to demonstrate that RATs are not caused by local regions of technical variation between replicates. Scale for panels **A** and **B**: 0–5 normalized replication signal. (**C**) Boxplots representing the distribution of RAT sizes in the three categories: Later-to-Earlier, Earlier-to-Later, and a subset of Earlier-to-Later RATs found in functional centromeres (CEN) [46]. Boxplot whiskers represent 1.5 x interquartile range (IQR). The axis is broken to show two values that are much higher than the others and correspond to large RATs in CEN 9 and CEN 10. However, it is important to note that the sizes of CEN RATs are underestimated, because centromeres contain variable numbers and sizes of blacklist regions, which break up what would probably be long continuous RATs (see Fig 4 and Table 2).

https://doi.org/10.1371/journal.pgen.1008623.g002

Robust RATs fall into two categories, those in which the strongest replication signal occurs later in the mitotic cycle than it does in the endocycle ("Later-to-Earlier" shift), and those in which the strongest replication signal occurs earlier in the mitotic cycle than in the endocycle ("Earlier-to-Later" shift). In addition, we separately characterized a subset of the Earlier-to-Later RATs that are located in functional centromeres ("Earlier-to-Later-CEN") using centromere (CEN) coordinates from [46]. Our stringent criteria identified RATs comprising only about 2% of the maize genome (Table 1), with 233 of the 274 total regions (representing 1.6% of the genome) in the Earlier-to-Later category. Non-CEN Later-to-Earlier and Earlier-to-Later RATs have similar size distributions, with median sizes of 141 and 135 kb, respectively (Fig 2C and Table 1). All of the CEN RATs fall into the Earlier-to-Later category and have a median size of 132 kb, similar to the non-CEN RATs. It is important to note, however, that the sizes of CEN RATs are underestimated because of numerous blacklist regions within the centromeres that break what are likely continuous RATs into several smaller parts in our analysis. Even though maize centromeres are remarkably well sequenced [46], they still contain some gaps and regions where reads cannot be uniquely mapped in the current B73 RefGen_v4 genome assembly. To account for this, we calculated an upper estimate of continuous RAT size by incorporating the coverage from the multiple RATs called within each centromere as

Table 1. RAT summary table.

RAT category	Count	Median size (kb)	Coverage (kb)	% of genome
Non-centromere				
Later-to-Earlier	41	141	6,291	0.30
Earlier-to-Later	192	135	26,907	1.28
Centromere				
Earlier-to-Later-CEN	41	132	7,668	0.36
Earlier-to-Later total	233	135	34,575	1.64
RAT total	274	135	40,866	1.94

A summary of the region count, median size, total genome coverage, and percentage of the entire genome represented in each RAT category.

https://doi.org/10.1371/journal.pgen.1008623.t001

well as coverage from RATs that extend past the previously determined CEN boundaries ("presumed CEN RATs"), and including the interspersed blacklist regions. This upper estimate of CEN RAT size, which is addressed in more detail below, ranges from about 0.9–2 Mb in each CEN (Table 2).

Non-centromeric RATs

We analyzed the non-CEN RATs for the content of genes and TEs, as well as the presence of histone modifications and functional annotations related to the genes within RATs. To assess whether the percentage of RATs containing genes differed from random expectation, we randomly shuffled coordinates corresponding to the non-CEN Later-to-Earlier and Earlier-to-

Table 2. Centromere RAT coverage and size estimates.

CEN	CEN size total (kb)[a]	CEN excluding blacklist (kb)[b]	≥10% DRT coverage (kb)[c]	Final RATs coverage (kb)[d]	Presumed CEN RAT coverage (kb)[e]	RAT count[f]	Continuous RAT size (kb)[g]	% CEN covered by continuous RAT[h]
1	350	329	15	-	-	-	-	-
2	1980	1275	1139	986	220	7	2037	86.9
3	1150	667	498	453	-	6	897	78.0
4	1430	968	706	538	14	7	945	65.1
5	2280	1806	1440	1407	-	7	1782	78.2
6	800	764	390	-	-	-	-	-
7	300	285	45	-	-	-	-	-
8	1540	1210	1190	1046	10	7	1419	91.5
9a	1650	1602	1510	1510	71	4	1641	95.7
9b	400	334	18	-	-	-	-	-
10	1390	1306	1269	1221	192	3	1506	94.5

a Centromere size reported in [46], which includes unmappable regions of known size.

b Centromere size, excluding blacklist regions (unmappable and multi-mapping) that are >6 kb in size. A gap distance of 6 kb was allowed during RAT analysis.

c The combined coverage of regions with ≥10% DRT called within each centromere, not including blacklist regions >6 kb.

d The combined coverage of final RATs called within each centromere, not including blacklist regions >6 kb. Final RATs contained at least one core region with ≥25% DRT (see Methods).

e In some centromeres, the called RAT extends past the previously reported CEN boundary. These areas are labeled 'presumed CEN RAT'.

f The number of individual RATs called in each centromere.

g Continuous RAT size is defined by summing the coverage of CEN RATs, presumed CEN RATs, and the interspersed blacklist regions to generate an upper estimate of full RAT size. This estimate still cannot take into account gaps in the centromere assemblies that are of unknown size.

h The percentage of the CEN size total that is covered by the estimated continuous RAT, without taking into account presumed CEN RATs outside CEN boundary.

A summary of the various estimates of coverage in individual centromeres of regions with DRT, and continuous RAT size.
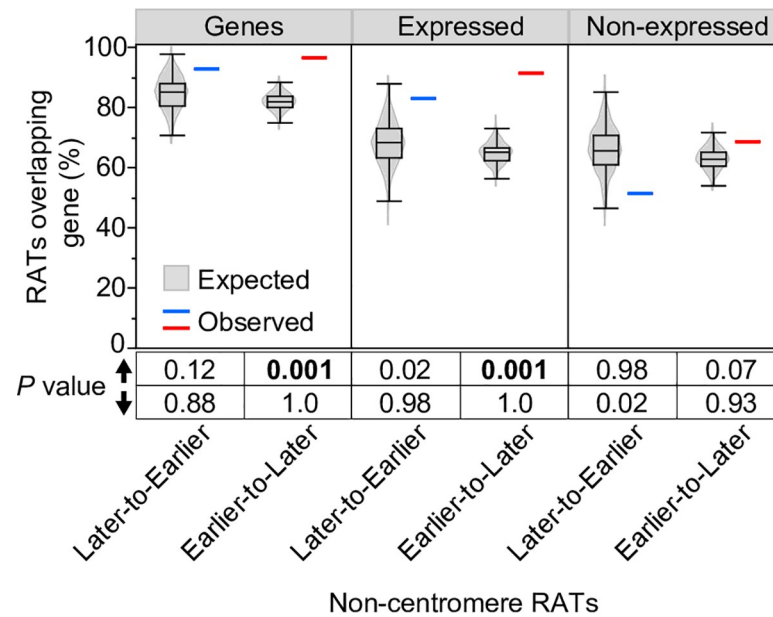
https://doi.org/10.1371/journal.pgen.1008623.t002

Fig 3. Permutation analysis of the percentage overlap of non-CEN RATs and genes. The percentage of RATs that overlap genes, expressed genes or non-expressed genes was calculated for non-CEN RATS and corresponding 1000 randomly shuffled sets (see Methods). The observed percentage for Later-to-Earlier (blue line) and Earlier-to-Later (red line) RATs are plotted alongside the expected percentage distribution of the 1000 random sets (grey violin plots overlaid with boxplots). Permutation *P* values below the graph were calculated from the proportion of the 1000 random sets that had a percent overlap value greater than (up arrow) or less than (down arrow) the observed value. Permutation *P* values ≤ 0.001 are considered evidence that the observed percent overlap is significantly different than random expectation.

Later RATs around the genome 1000 times and calculated the percentage of randomly shuffled regions that overlap genes in each of the 1000 sets. These "expected" random distributions were compared to the observed percent overlap values found for Later-to-Earlier and Earlier-to-Later RATs and a permutation *P* value was calculated (see Methods). We found that 93% and 96% of Later-to-Earlier and Earlier-to-Later RATs, respectively, contain at least one annotated gene and usually contain a small cluster of, on average, 2–3 genes (Fig 3 and S3 Table). The observed 96% percent overlap of Earlier-to-Later RATs with genes is significantly greater than expected by chance (permutation *P* value = 0.001; Fig 3). To assess which of these genes are expressed, we used root-tip RNA-seq data that are not specific to mitotic or endocycle cells, and found that although only 50% of the 682 genes found in non-CEN RATs are expressed at a meaningful level (FPKM ≥ 1), 83% and 91% of Later-to-Earlier and Earlier-to-Later RATs, respectively, still contain at least one expressed gene (S3 Table). The observed 91% overlap of Earlier-to-Later RATs with expressed genes is also significantly greater than expected by chance (permutation *P* value = 0.001; Fig 3). In contrast, we found only 51% and 68% of Later-to-Earlier and Earlier-to-Later RATs, respectively, contain genes not expressed in the root (Fig 3). These values for non-expressed genes are not significantly different from random expectation, indicating that the enrichment of genes in Earlier-to-Later RATs is mainly driven by the expressed genes.

The significant enrichment of genes expressed in the root in Earlier-to-Later non-CEN RATs suggests the possibility that these regions may be related to shifts in gene expression. However, we were unable to directly compare expression of genes in RATs in mitotic and endocycling cells because we could not obtain RNA of sufficient quality to sequence from fixed, sorted nuclei. Instead, we assessed a selection of gene-associated histone post-

translational modifications in sorted non S-phase 2C, 4C and 8C nuclei. In our previous work in maize root mitotic cells, we showed that trimethylation of H3 lysine 4 (H3K4me3) and acetylation of H3 lysine 56 (H3K56ac) modifications tend to colocalize on active genes and are associated with earlier replicating regions, while trimethylation of H3 lysine 27 (H3K27me3) tends to be on repressed genes regardless of their RT [19]. For each ploidy level, we quantified the percentage of genes within RATs that have each mark, as well as the fold enrichment relative to input for called peaks within genes. There are very few differences between ploidy levels in the number of genes bearing these marks (S7D Fig), but there are some minor shifts in the peak enrichment in 8C nuclei compared to 2C (S7A–S7C Fig). The clearest shift is a decrease in H3K4me3 enrichment found on expressed genes in Earlier-to-Later RATs (S7B Fig), which suggests these genes may have decreased expression in endocycling cells.

We also performed a gene ontology (GO) analysis for the genes found in non-CEN RATs to ask if there are functional annotations enriched in genes that shift RT. For this analysis, we focused on the genes that we identified as expressed in the root tip (S2 Spreadsheet). We found 44 significantly enriched GO terms for genes within Earlier-to-Later RATs, including biological process and molecular function terms related to gene expression, DNA/RNA metabolism, and the cell cycle (S8 Fig). A wide variety of significant cellular component GO terms were also found, which may relate to various differentiation processes occurring in endocycling cells. There are no significant GO terms for genes within Later-to-Earlier RATs, though the presence of only 52 expressed genes in this RAT category made it difficult to fully assess significance. Taken together, these analyses of transcription-related histone modifications and functional annotations suggest a role for gene expression changes in the Earlier-to-Later RATs. Given that these regions are shifting to a later RT in the endocycle, a decrease in gene expression would be expected [19]. However, more work will be needed to confirm this hypothesis.

The general organization of the maize genome is genes clustered in "islands" interspersed with blocks of transposable elements [50–52]. We used a permutation strategy similar to that described above to estimate the significance of any differences in percent coverage of TEs and individual TE superfamilies in non-CEN RATs. Observed coverage values were compared to random expectation, estimated from 1000 randomly shuffled sets. The TE annotations were from the recent B73 RefGen_v4 TEv2 disjoined annotation, where every bp is assigned to a single TE [47]. We found no TE superfamilies with percent coverage values in non-CEN RATs that are significantly different from random expectation at the permutation P value threshold of 0.001, although one TE superfamily, RLG/Gypsy, is very close to the threshold for being called significantly depleted in Earlier-to-Later RATs (permutation P value = 0.002; S9 Fig). RLG/Gypsy elements make up 32% of the coverage of Earlier-to-Later RATs, and from 32–40% of the coverage of the randomly shuffled sets. We also found that the percent AT content in RATs is similar to that of the genome as a whole, with median values of 55% and 56% for Later-to-Earlier and Earlier-to-Later RATs, respectively, compared to a median value of 55% for the whole genome (S10 Fig). Taken together, there is no evidence of a difference in AT content or a major enrichment or depletion in the coverage of specific TE superfamilies in non-CEN RATs, with the possible exception of a minor depletion of RLG/Gypsy element coverage in Earlier-to-Later RATs. Given that RLG/Gypsy elements have by far the most abundant coverage across the B73 reference genome [46], this minor depletion may be related to the increased presence of genes in Earlier-to-Later RATs (Fig 3).

Centromeric RATs

Functional centromeres are defined by their content of nucleosomes containing the centromere-specific histone variant known as CENH3 in plants and CENP-A in animals. CENH3/

CENP-A makes up only a small percentage of the total H3 population in centromeres, but plays an important role in recruiting kinetochore proteins [53–55]. Maize is unusual among higher eukaryotes in that a majority of centromeric reads can be uniquely mapped [56]. In our Repli-seq data, for example, we found that on average 45% of all reads that map to centromeres could be uniquely mapped to a single location (S11 Fig). Only these uniquely mapping reads were used for further analysis. In addition, most of the maize centromere assemblies are relatively intact, and functional centromeres have been located by mapping ChIP-seq reads for CENH3 [46]. When combined with our replication timing data, these features of the maize system create a unique opportunity to assess RT programs for centromeres.

Our analysis found large, robust RATs across seven of the ten centromeres (Figs 4C, 4D and S12). These seven centromeres (CEN 2, 3, 4, 5, 8, 9 and 10) were previously classified as "complex" because they contain a mixture of retrotransposons with some centromere satellite repeat arrays (*CentC*; [56, 57]). In the RefGen_v4 genome assembly, CEN 9 has two called CENH3-binding regions [46], which we refer to as CEN 9a and 9b (Fig 4C; grey bars). Interestingly, we only found robust RATs in the larger CEN 9a, with the smaller CEN 9b showing very little RT shift (Fig 4C and Table 2). The cumulative coverage of RATs in each complex CEN ranges from 0.45–1.5 Mb (Table 2). However, because each centromere includes blacklist regions that vary in size and number, automated analysis did not identify the true sizes of the RATs. Therefore, we also calculated an upper estimate of continuous RAT size in each centromere, including the multiple RATs called by automated analysis as well as the interspersed blacklist regions. We also included coverage from CEN RATs that extend somewhat past the previously determined CEN boundaries ("presumed CEN RAT"). This upper estimate ranges from about 0.9–2 Mb in different centromeres, and covers between 65–95% of the previously defined functional centromere for the seven complex CENs (Table 2).

Upon inspecting the RT profiles for each of the complex CENs we found the strongest replication signals occurring mainly in mid S in mitotic cells, but changing to primarily late S in endocycling cells. It is also noteworthy that though replication occurs mainly in mid S in mitotic cells, there are some distinct peaks of early replication inside or directly adjacent to the called centromere (indicated by black arrowheads in Figs 4 and S12) in all but one (CEN 5) of the complex centromeres. These early peaks remain in the endocycle, though usually there is some reduction in early signal with a concomitant increase in mid signal at the same location.

The remaining three centromeres (CEN 1, 6, and 7) were previously characterized as "simple" because they mainly contain large arrays of the *CentC* repeat [56, 57]. In our analysis, the simple centromeres showed, at most, small RT shifts that did not meet our criteria for a robust RAT (Table 2, Figs 4A, 4B and S12). However, *CentC* repeats are not well represented in the reference genome assembly, so it is not possible to analyze RT profiles for the complete simple centromeres. Portions of CEN 7 that are present in the assembly replicate mainly in mid S phase in both mitotic and endocycling cells (S12 Fig), while sequences in the assemblies for CEN 1 and CEN 6 are mostly late replicating in both types of cells, with some minor RT changes across small regions (Fig 4A and 4B and Table 2).

Because of the issues with computationally identifying continuous RATs we chose to focus the following set of analyses on the entire CENH3-binding region of each chromosome (excluding blacklist regions). We calculated the difference in early, mid and late replication signal (endocycle minus mitotic) from RT profiles by averaging across 100-kb static windows. For comparison, we also calculated the replication signal differences in pericentromeres, which were arbitrarily defined as the ± 1 Mb flanking the CENH3 region. We inspected all RT differences in the centromeres and pericentromeres by not requiring that the DRT be compensated by an opposite shift in the other S-phase fractions. Early and mid replication signals across the complex centromeres decrease and late replication signals increase in endocycling
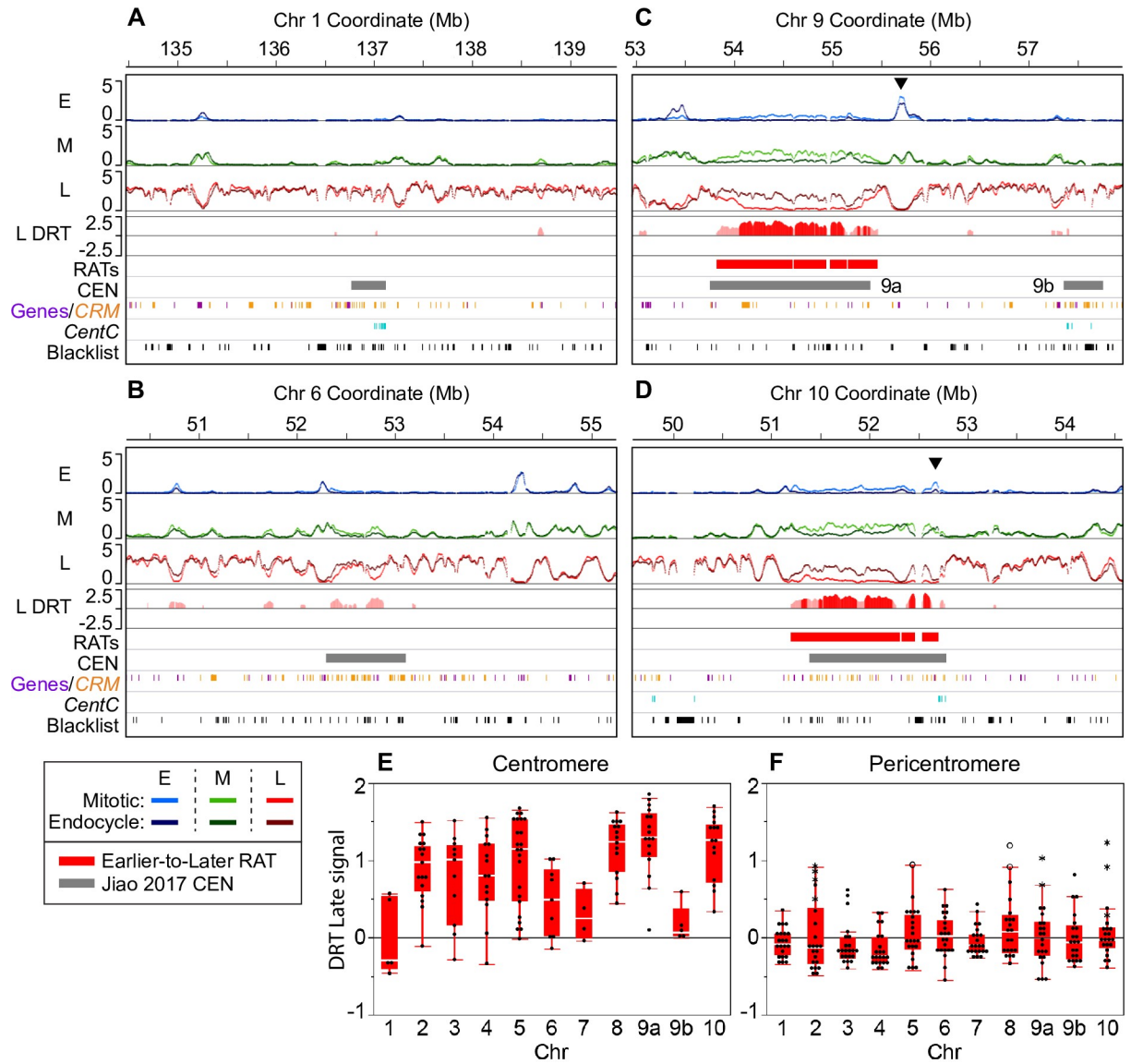
Fig 4. Large RATs correspond to functional centromeres. Our analysis found large RATs, sometimes broken by blacklist regions (black tick marks at the bottom of each panel) at each of the seven "complex" maize centromeres. The remaining three "simple" centromeres (on chromosomes 1, 6, and 7) showed various levels of DRT that did not meet the criteria for calling RATs in our initial analysis. (**A–D**) Each 5-Mb region shown contains early (E), mid (M) and late (L) RT profiles with mitotic and endocycle data overlaid (scale: 0–5 normalized replication signal). The difference in late replication signal (endocycle minus mitotic; labeled "L DRT") for windows where the difference was compensated by an equal and opposite difference in the early and/or mid profiles is also shown. Late replication signal differences compensated at the $\geq$ 10% threshold (light red), and those compensated at the $\geq$ 25% threshold (dark red) are shown, but only regions that contained at least one $\geq$ 25% shift were classified as robust RATs in our initial analysis. Two examples of simple centromeres, CEN 1 (**A**) and CEN 6 (**B**), and two examples of complex centromeres, CEN 9 (**C**) and CEN 10 (**D**) are presented. The black arrowheads in panels **C** and **D** denote example regions with a peak of early replication signal within or adjacent to the centromere that also shows an increase in mid replication signal in the endocycle (for other examples, see S12 Fig). Colored boxes below the RT profiles denote Earlier-to-Later RATs (red) and the functional centromere (grey; [46]). Chromosome 9 contains two called CEN regions labeled 9a and 9b. The colored tick marks correspond to elements of centromeric retrotransposons of maize (*CRM*) families 1–4 (orange; [47]), gene annotations (purple; [46]), and mappable *CentC* satellite repeats (teal; [57]). Blacklist regions are indicated by black tick marks in the lowest track. (**E** and **F**) DRT (endocycle—mitotic) between late RT profiles for each centromere (**E**) and corresponding pericentromere (**F**; ± 1 Mb) were calculated in 100-kb static windows. In panel **F**, asterisks indicate DRT values from windows where an Earlier-to-Later-CEN RAT extends past the called CEN boundary [46] into the pericentromere (also see Table 2); open circles indicate windows that contain a non-CEN Earlier-to-Later RAT that met our compensation criteria. DRT values between early and mid profiles are shown in S13 Fig.

https://doi.org/10.1371/journal.pgen.1008623.g004

cells, reflecting a large shift toward late replication. The DRT values for the late profile in centromeres and pericentromeres are shown in Fig 4E and 4F, respectively, while the DRT values for early and mid profiles are shown in S13 Fig. Interestingly, the DRT tapers off towards the edges of the functional centromere (see profiles in Figs 4C, 4D and S12), and there is striking congruity in the replication signals for mitotic and endocycling cells in the immediately adjacent pericentromere regions (Fig 4A–4D). The few RT shifts in pericentromeric regions are smaller in size and much less dramatic than those in the centromere proper (Fig 4F). Moreover, very few (8%) of pericentromeric windows with DRT are compensated by an equal and opposite shift in the other S-phase profiles (S4 Table), suggesting many of these uncompensated differences may result from technical variation rather than from meaningful biological differences. In contrast, nearly all (85%) of the centromeric windows have compensated RT shifts.

Genomic elements and features in centromeres

Maize centromeres contain varying amounts of tandemly arrayed *CentC* repeats (single repeats of 156 bp in length; [41]) as well as several *CRM* retrotransposon families interspersed with elements from a few other retrotransposon families [44, 52, 58, 59]. *CentC* repeats and *CRM* elements are also present in the adjacent pericentromeres where there is no CENH3 binding [52, 58]. In RefGen_v4, there are also fifty annotated genes within centromeres. We asked if all of these sequence elements in centromeres behave similarly in the mitotic to endocycle transition, or if certain elements show larger RT shifts than others. We also asked if all three types of sequence elements show similar RT changes in centromeres versus pericentromeres. Given that the replication signal values were aggregated in 3-kb windows, we only included elements that covered at least half a window (1.5 kb) in our analysis. Fig 5 summarizes data on these questions for the complex centromeres, while data for the simple centromeres are shown in S14 Fig. Similar results were found when all elements were included (S14 Fig).

The results for the two dominant *CRM* families, *CRM1* and *CRM2*, are similar (S14 Fig), so these families were grouped together in Fig 5C. When present in centromeres, all three major classes of elements–genes, *CRM1/2*, and *CentC* repeats–clearly replicate later during the endocycle than in the mitotic cycle (Fig 5). In contrast, genes and *CRM* elements in the pericentromere show little or no timing shifts. A full analysis of the replication times of *CentC* repeats in pericentromeres is hampered by the limited representation of this repeat class in the genome assembly (Figs 5D and S14E).

Chromatin features in centromeres

We also examined activating (H3K56ac and H3K4me3) and repressive (H3K27me3) histone H3 post-translational modifications to look for epigenetic changes in centromeres after endocycle replication. It was previously reported that some H3K4me3 and H3K27me3 peaks of enrichment occur in the centromere, mainly associated with genes [60]. We asked whether genes that have these modifications continue to have them after mitotic and endocycle replication, and found very few changes in the number of genes with these modifications at each ploidy level (S15 Fig). There was also very little change in the fold enrichment of these histone marks in centromere genes when comparing 2C, 4C and 8C nuclei.

We also investigated the levels of dimethylation of histone H3 lysine 9 (H3K9me2) enrichment in each centromere after mitotic and endocycle replication. Previous work indicated there is a depletion of H3K9me2 in centromeres relative to adjacent pericentromeres [61, 62], which we observed as well (S16 Fig). Traditional peak calling tools are not effective for H3K9me2 because of its even distribution across the maize genome. Instead, we estimated the
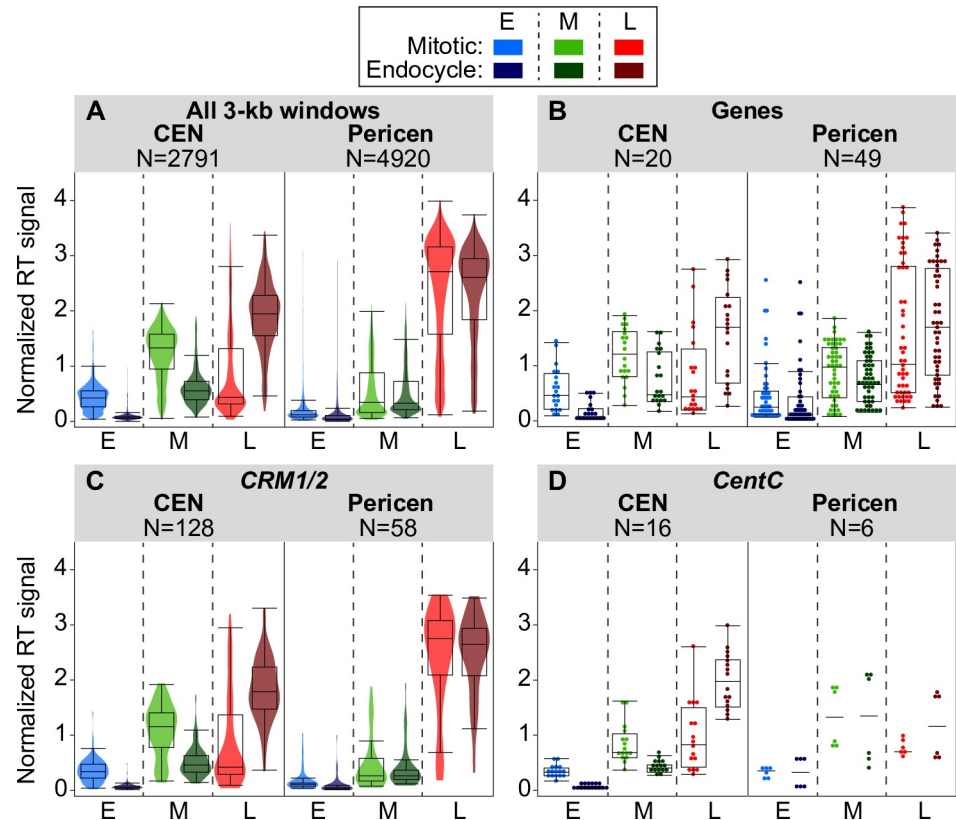
Fig 5. Comparing replication times for genomic features in complex centromeres and corresponding pericentromeres. (**A–D**) Boxplots comparing replication signals during mitotic and endocycle S phases for centromeres, pericentromeres (± 1 Mb), and genomic features within them. The panels show the distributions of replication signals in early (E), mid (M), and late (L) S for all 3-kb windows (**A**), annotated genes (**B**), *CRM1/2* elements (**C**), and mapped *CentC* repeats (**D**) in centromeres and pericentromeres. For panels **A** and **C**, colored violin plots are overlaid, while for panels **B** and **D**, individual data points are shown because of the smaller number of data points. The number of windows or elements included in each analysis is indicated above each graph. Only elements that covered at least 50% of a 3-kb window were included in each analysis, though results were similar when all elements were included (S14 Fig). Boxplots for all elements in simple centromeres, as well as for the individual *CRM1* and *CRM2* families are in S14 Fig.

https://doi.org/10.1371/journal.pgen.1008623.g005

fold enrichment relative to the corresponding DNA input control by calculating the percent of total H3K9me2 ChIP reads in a given centromere region (using coordinates from [46]) and dividing by the percent of total input reads corresponding to that centromere in three biological replicates. We found a similar H3K9me2 average depletion for all centromeres and for 2C, 4C and 8C nuclei relative to their corresponding input (averages for individual CENs 0.75–0.88, 0.83–0.91 and 0.85–0.93 for 2C, 4C and 8C nuclei, respectively), although values for 4C and 8C nuclei were consistently slightly higher than those for 2C nuclei (S16A Fig). The maize CENH3 N-terminal tail lacks some of the conserved motifs found in canonical histone H3 (see S3 Table in [63]), so H3K9me2 enrichment is likely to occur in the interspersed H3 nucleosomes. Altogether, we found very little change in centromeres between non S-phase 2C, 4C and 8C nuclei in the histone H3 post-translational modifications we assessed by ChIP-seq.

Centromeric histone H3 in mitotic and endocycling centromeres

Unlike the canonical histone H3, CENH3 is not replaced in a replication dependent manner in higher eukaryotes, resulting in a dilution of CENH3 relative to centromeric DNA during S

phase [64, 65]. New CENH3 is incorporated into nucleosomes after the completion of S phase, but the timing of its integration into centromeric chromatin differs for plants, flies and humans (reviewed in [66]). In the plants tested thus far, deposition of CENH3 has been reported to occur between late G2 and metaphase [67–70].

Because mitosis does not occur in the endocycle and centromere function is presumably not required, we speculated that CENH3 might remain at low levels following DNA replication in endocycling cells. This hypothesis is supported by cytological studies of *Arabidopsis* endo-polyploid nuclei showing the CENH3 signal does not increase in parallel with the total DNA content or the signal for 180-bp centromeric repeats [68, 69]. To test this hypothesis with maize centromeres, we used a maize anti-CENH3 antibody [58] for ChIP-seq analysis of CENH3 binding in sorted non S-phase 2C, 4C and 8C populations of nuclei. It is important to note that the 4C nuclei come from a mixture of cells, some of which will return to the mitotic cycle, while others will remain at 4C, and still others will continue on to the endocycle (at least 15% of nuclei in the combined 0–3 mm region). We asked whether the location or level of CENH3 enrichment changed after DNA replication in the mitotic cycle or the endocycle. For visualization of CENH3 localization, ChIP-seq uniquely mapping read counts from three biological replicates for each ploidy level were aggregated in 3-kb windows and normalized to the level of a uniform 1× genome coverage, so that corresponding windows in the different ploidy level profiles were comparable. The normalized read count in each 3-kb window was then divided by the normalized read count for input DNA of the corresponding ploidy to calculate a fold enrichment value for CENH3 binding sequences in that window. The spatial distribution of CENH3 enrichment across the centromeres remained the same in 2C, 4C, and 8C cells. This is illustrated for CEN 9 and CEN 10 in Fig 6A and 6B, and data for the rest of the centromeres are shown in S17 Fig. There are also a few small spikes of CENH3 enrichment outside the called centromere (e.g. seen in Fig 6 and S17 Fig, but also occasionally further out on the arms). These spikes also remain in the same location between 2C, 4C and 8C cells. Some of them could be related to misassembly of the reference genome. However, if real, these ectopic CENH3 peaks are less numerous and more persistent in G2 (4C) than those recently observed in HeLa cells [71].

To compare total CENH3 content of entire centromeres at different ploidy levels, we calculated the percent of total CENH3 reads found in a given centromere and made a ratio to the percent of total reads from the corresponding input DNA in that centromere separately for each biological replicate, as described above for H3K9me2. The CENH3 average fold enrichment relative to total DNA content is similar for 2C and 4C nuclei in each of the complex centromeres (Fig 6C), with an average 4C/2C enrichment ratio of 1.1 (S5 Table). However, CENH3 enrichment decreases with the increase in ploidy from 4C to 8C (Fig 6C). As noted above, the 4C nuclei come from a mixed population of cells, only a fraction (ca. 15% in the combined 0–3 mm root region) of which will enter the endocycle. Because of this ambiguity, we chose to focus on the CENH3 enrichment in 8C nuclei. In these nuclei, we found, an average 8C to 2C enrichment ratio of only 0.7 (S5 Table). CENH3 enrichment values for simple centromeres were lower and slightly more variable, likely because of assembly issues. In both cases, however, the ratio of CENH3 enrichment in 8C cells to that in 2C cells is higher than 0.5, the ratio that would be expected if there was no incorporation of new CENH3 after endocycle replication, but smaller than the 1.0 ratio expected if there was full replacement (S5 Table). It is worth noting that these data refer to post-replication 8C nuclei, which exited S phase prior to the time of analysis, and that post-replication 4C nuclei show no dilution of CENH3 relative to DNA content. Thus, our data are consistent with a hypothesis in which the average CENH3 to DNA ratio in the 8C population is only partially restored after completion of S phase.
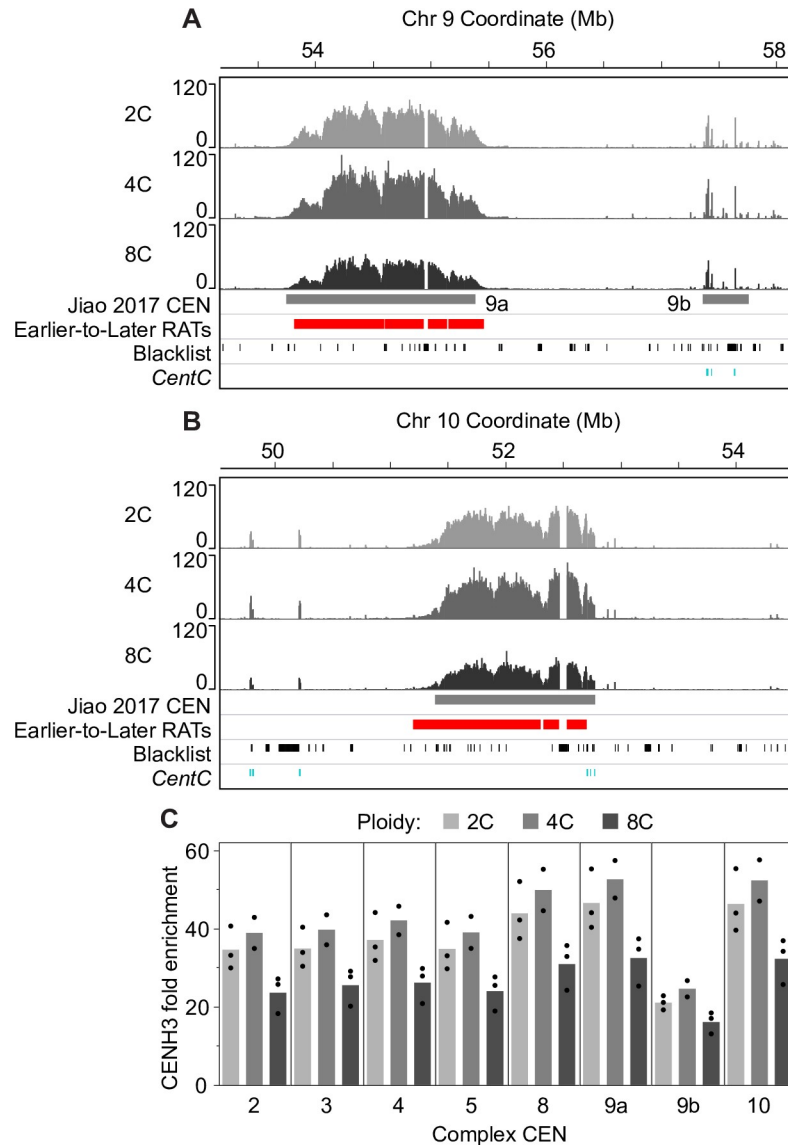
Fig 6. CENH3 localization and enrichment in mitotic and endocycling centromeres. We profiled CENH3 binding by ChIP-seq in flow sorted, non S-phase nuclei with 2C (before mitotic replication), 4C (after mitotic replication) and 8C (after endocycle replication) DNA contents. (**A** and **B**) CENH3 localization patterns for 2C, 4C and 8C nuclei in CEN 9a and 9b (**A**) and CEN 10 (**B**). Scale in both panels is 0–120 fold CENH3 enrichment relative to input. Colored boxes below the CENH3 profiles denote the previously identified functional centromere (grey; [46]), and Earlier-to-Later-CEN RATs (red). Tick marks in the bottom two tracks indicate blacklist regions (black) and mapped *CentC* repeats (teal). (**C**) We used the ChIP-seq datasets from 2C, 4C and 8C nuclei to estimate the CENH3 average fold enrichment relative to DNA content for complex centromeres by calculating the percent of total CENH3 reads found in a given centromere (using coordinates from [46] and dividing by the percent of total input reads corresponding to that centromere. Black dots represent the individual values from biological replicates. Data for simple centromeres are shown in S17B Fig.

https://doi.org/10.1371/journal.pgen.1008623.g006

Discussion

The maize root tip includes a naturally occurring developmental gradient, with cells in the meristem region (ca. 0–1 mm) primarily undergoing mitotic cell cycles, while a subpopulation of cells in the transition zone (ca. 1–3 mm) enters a developmentally programmed endocycle prior to further differentiation [8, 9]. Even though endocycling is very common in plants and

plays essential roles in differentiation and the development of specialized tissues, cell size increases, and stress responses [2, 5, 72, 73], replication timing (RT) programs have not yet been characterized for alternative cell cycles, such as the endocycle.

We generated whole genome Repli-seq data for root cell nuclei undergoing DNA replication in either the mitotic cycle or the endocycle, making use of *in vivo* EdU labeling of intact root tips and two-color fluorescence activated nuclei sorting. By doing so, we avoided potential artefacts caused by cell synchronization [74] and chromosome aberrations often found in plant and animal cell cultures (e.g. [75–77]). We present replication timing activity profiles (RT profiles) for early, mid and late replication separately, instead of collapsing the data into an early:late ratio as many studies do. The rationale for this approach is that, for roughly one third of the maize genome, we previously found heterogeneity in mitotic RT–e.g. regions of the genome in which root tip cells exhibit significant replication activity in both early and mid S, or both mid and late S [19]. An additional advantage to presenting the RT profiles separately is the ability to assess whether there are concomitant or "compensated" changes in a region at multiple stages of S phase. This compensation criterion helped us separate RT shifts that could be subject to technical error, such as alterations in flow sorting gates, from shifts that are more likely to represent meaningful changes in the population preference to replicate a replicon or cluster of replicons at a particular time in S phase.

The current study sought to investigate whether the mitotic RT program is maintained as endocycling cells transition from 4C to 8C, despite the need to replicate twice as much DNA and the initiation of various root cell differentiation pathways. Extending our previous cytological observation that spatiotemporal patterns of replication are similar in mitotic and endocycling cells [22], we found that RT programs at the sequence level are also very similar. Pearson's correlation coefficient values comparing data from the two types of cell cycles were nearly identical to those for biological replicates within each type. The high level of similarity is particularly noteworthy in the case of the early RT profiles, given that the flow sorting gate for early replicating nuclei in the endocycle had to be adjusted to minimize contamination from late replicating mitotic nuclei (Fig 1C). This overall conservation of the RT program in the endocycle is consistent with a recent study in *Drosophila* follicle cells which found no regions of differential RT between mitotic and endocycling cells [78]. Additionally, the global maintenance of RT in the two types of cell cycles in maize roots suggests that the process of re-establishing the RT program must be similar in both. In animal systems, re-establishment of the RT program has been shown to occur in G1 of each cell cycle at a "timing decision point" [79], however the details of this process have not been studied in plants.

Most plants fully replicate their genome during endocycles [80], although there are a few exceptions (e.g. various orchid species; [31, 32]). We found very little evidence for over- or under-replication occurring in endocycling maize root cells, unlike the distinctive over- and under-replication found in *Drosophila* endocycles (reviewed in [27] and references therein). Our result is consistent with earlier cytological reports that whole chromosomes, as well as repetitive knobs and centromeres, are completely replicated in the highly endopolyploid maize endosperm [33].

In contrast to the global maintenance of the RT program, we observed a small fraction of the maize genome that exhibits some difference in RT (DRT) between the two types of cell cycles. Approximately 11% of the genome showed compensated DRT at a stringency level of $\geq 10\%$ difference in replication signal (see Methods). However, with the notable exception of centromeric regions, which are discussed in more detail below, we chose to characterize only the most robust Regions of Altered Timing (RATs), defined by the criteria of containing a core region with compensated DRT at a stringency level of $\geq 25\%$ difference in replication signal. These robust non-centromeric RATs comprise only 1.6% of the genome (centromeric

RATs comprise an additional 0.4% of the genome), and the size range of individual non-centromeric RATs (39–387 kb, median 138 kb) is consistent with our previous observation that regions of coordinate replication in maize are ~50–300 kb in size [19]. This may include from one to a few replicons, based on previous estimates of replicon size in monocot plants [49].

The first 1-mm of the maize root contains the meristem and precursors for at least ten different cell types. Only some of these cell types enter the endocycle prior to cell elongation [9]. If there are differences in the RT programs of different cell types, some or all of the non-centromeric RATs may be associated with shifts in the relative contribution of different cell types to the two samples of nuclei, rather than to endocycling *per se*. Research in metazoans has revealed ~8–20% of their genomes can shift RT between cell types [11–13, 81–83]. In mammals, these RT shifts generally involve large regions or "domains" in the megabase size range (reviewed in [26]). These RT domains are much larger than the non-centromeric RATs in maize, even though the maize genome is similar in size to the human and mouse genomes. However, in the much smaller *Drosophila* genome, regions that show RT shifts between cell types are more similar in size to the maize non-centromeric RATs [81, 83].

The vast majority of the non-centromeric RATs involved RT shifts from Earlier-to-Later, with a significant enrichment for genes expressed in the root tip, but not for non-expressed genes (FPKM < 1). This result suggests the possibility that RT shifts may be related to shifts in gene expression. Unfortunately, we have been unable to follow transcriptional changes in endocycling nuclei directly, as we have so far not been able to isolate RNA of sufficient quality to characterize transcripts from fixed, sorted nuclei. However, our analysis of activating and repressive histone modifications uncovered only minor changes in the enrichment and location of these marks within RAT genes after endocycle replication. The lack of notable changes in the proportion of RAT genes bearing H3K56ac and H3K4me3 modifications after the endocycle suggests that these histone marks are permissive to changes in RT. Nonetheless, the direction of the change in H3K4me3 enrichment on genes in Earlier-to-Later RATs after endocycle replication (S7B Fig) is consistent with the hypothesis that a shift to later RT may accompany a decrease in gene expression. Many studies have identified a correlation between RT and transcriptional activity (reviewed in [26]), but there are also multiple examples of these processes being uncoupled (e.g. [14, 84]).

In the case of centromeres, it is easy to imagine that the large shifts to later replication are related specifically to endocycling, because endocycling cells presumably no longer require functional centromeres. Though often broken by unmappable and multi-mapping ("blacklist") regions in the genome assembly, centromeric RATs when combined across blacklist regions are much larger in size (0.9–2 Mb) than the non-centromeric RATs and cover the majority of each of the seven complex centromeres (Table 2). These seven centromeres, which are well assembled in the maize B73 RefGen_v4 genome, contain satellite repeats interspersed with retrotransposons [46, 56], enabling almost 50% of our sequencing reads that map to these centromeres to be uniquely positioned. In most species, in which centromeres contain large numbers of tandemly arrayed satellite repeats, it is difficult to map centromeric sequence reads to unique positions and, thus, to fully assess centromeric RT patterns [85]. Though yeast centromeres replicate in early S phase [86–89], most higher eukaryotes replicate centromeres asynchronously through mid to late S [64, 90–95]. Many of the reports in higher eukaryotes are based on cytological observations, membrane hybridization, or PCR data with limited resolution. Even a recent genomic analysis of centromeric RT in human cell lines was significantly limited by the quality of the human centromere assemblies, and could only uniquely map ~15% of centromeric reads [85]. Centromere replication in plant species, assessed mostly by cytological methods, has variously been reported to occur in early, mid or late S [96–99], though it is often unclear if the analysis was of sufficient resolution to distinguish the RT of

centromeres from that of adjacent pericentromeres. In contrast, we have provided a high-reso-
lution analysis of the distribution of replication times across maize centromeres, and com-
pared RT of centromeres to adjacent pericentromeres.

These analyses revealed several features shared by the RT programs of the seven complex
maize centromeres. For example, in mitotic cells there are a few distinct peaks of early replica-
tion (e.g. arrowheads in Fig 4 and S12 Fig), flanked by mainly mid replication activity that is,
in turn, flanked by regions of late replication at the edges of the functional centromere. Except
for these few early regions, which show an increase in mid replication activity in the endocycle,
entire centromeres and the genes, retroelements and *CentC* repeats within them–undergo a
shift to late replication in the endocycle. As a result, the RT of the complex centromeres in the
endocycle becomes much more similar to that of the immediately adjacent pericentromeric
regions, which replicate primarily in late S phase in both mitotic and endocycling cells. Late
replication of pericentromeres is expected based on our previous cytological observations in
mitotic and endocycling nuclei [22] and the typical replication time of highly compacted het-
erochromatin in many systems.

The presence of distinct peaks of early replication in or adjacent to functional centromeres
(arrowheads in Fig 4 and S12 Fig) is noteworthy because they signify a population preference
for replication initiation in early S phase at these loci. This observation is of particular interest
because in the yeast *Candida albicans*, centromeres contain a replication origin that is the first
to initiate on its respective chromosome and plays a role in centromere specification [89].
Work in other yeasts has shown evidence that the phenomenon of early firing origins in and
near centromeres is present across a range of yeast phylogeny and that the presence and func-
tion of the centromere itself influences RT [100–102]. In maize, there is no evidence that these
early regions in and adjacent to centromeres are the first to replicate on the entire chromo-
some, but they are earlier replicating than their surroundings. Origin mapping experiments
(e.g. [103–105]) would be required to distinguish whether these early regions contain single or
small clusters of origins, and the location of any other origins in centromeres that may initiate
in mid or late S phase.

Unlike complex centromeres, the three simple centromeres of maize show less drastic RT
changes that occur over smaller regions. These simple centromeres are not as well assembled
as the complex centromeres [56, 57], and we cannot assess RT for the possibly large portions
of these centromeres not present in the genome assembly. One potential interpretation of our
results is that the simple centromeres have distinct RT programs that show less timing shift in
the endocycle, possibly related to their different sequence composition. Alternatively, the miss-
ing portions of the simple centromere assemblies could be replicating more like the complex
centromeres. Because simple centromeres are known to primarily contain large *CentC* arrays
[56, 57], the second hypothesis is supported by our analysis of mapped *CentC* satellite repeats
in all centromeres, which showed that, as a group, these repeats consistently shift RT from mid
to late. Another piece of evidence comes from our analysis of complex centromeres, which
showed that the magnitude of the RT change tapers off toward the outer edges of the func-
tional centromere. One can speculate that the simple centromere assemblies are comprised
mostly of the sequences at the edges of the actual centromere, which would still be anchored to
nonrepetitive regions in the genome assembly. As in complex centromeres, these edge
sequences might have a smaller RT shift than internal sequences. Future cytological experi-
ments, using a combination of flow sorted, EdU-labeled nuclei and techniques for identifying
maize chromosomes [106, 107] could help address questions related to the RT of simple
centromeres.

The centromere-specific histone variant, CENH3 (also called CENP-A in animal systems)
plays an important role in recruiting kinetochore proteins [53–55]. In metazoans, it has been

shown that CENP-A is distributed among sister centromeres during replication, but the full complement of new molecules is not redeposited until later [65, 108]. However, there are differences in the timing of deposition of CENH3/CENP-A among eukaryotes. Deposition occurs from S phase to G2 in yeasts, while in plants and protozoans it occurs from late G2 to metaphase, and in metazoans it occurs mostly during G1 (with the exception of some *Drosophila* cell types in metaphase to G1; reviewed in [55, 66, 70]). These interesting differences between phylogenetic groups in the timing of CENH3/CENP-A deposition suggest there may also be differences in the mechanisms and regulation of deposition that need to be explored further [69]. In our analysis of CENH3 enrichment relative to DNA content in maize root cells, the population of 4C nuclei appear to have a full complement of CENH3, which would be consistent with the previous results for plant species. This result supports a hypothesis that the subpopulation of 4C cells entering the endocycle also carry a full complement of CENH3. If this hypothesis is correct, our data for 8C nuclei imply that CENH3 is only partially replaced after DNA replication in the endocycle. Because the population of 8C nuclei we analyzed likely represents a mixture of cells that recently exited the endocycle S phase and others that exited some time ago we cannot determine whether CENH3 is fully restored in some cells and not others, or if it might be fully restored at a later time. However, our data suggest that the ratio of CENH3 to DNA is not immediately restored in the 8C population, and that the lower ratio is widely distributed across all ten centromeres.

It is unlikely that endocycling cells will ever re-enter the mitotic cycle [1, 109, 110], and it is not clear why endocycling cells would maintain or redeposit CENH3 nucleosomes at all unless CENH3 has roles outside of mitotic cell division. A recent study in *Drosophila* midgut cells found that CENP-A is required in post-mitotic and differentiated cells, and proposed that the loading of CENP-A in endocycling cells is essential for maintaining chromosome cohesion [111]. This possibility has not yet been tested.

Centromeres are considered to be epigenetically specified, as there are no unique sequences in the functional centromere that are not also found in the adjacent pericentromere (e.g. reviewed in [53, 112]). With this in mind, we tested whether changes in enrichment levels of CENH3 nucleosomes, or several modifications to canonical H3 nucleosomes, could explain the large shift to later replication of centromeres in endocycling cells. These studies only uncovered very small changes in activating and repressive histone H3 modifications in centromeres after endocycle replication. The magnitude of the change in CENH3, while somewhat larger, was not on the scale of the change in RT. It is possible that more significant changes might be found in epigenetic factors that we did not investigate, for example changes in DNA methylation patterns or other histone post-translational modifications. A variety of modifications to CENP-A nucleosomes have been identified, (reviewed in [113]), but very little is known about CENH3 modifications in plants [114, 115], highlighting an area for future research. Experiments in human cells identified cell cycle related interchanges of acetylation, monomethylation and ubiquitination at the lysine 124 residue of CENP-A [116, 117]. Mutations of this residue led to replication defects and alterations to centromeric RT [117].

An interesting question for future investigation is whether changes in chromatin conformation or 3D positioning in the nucleus are associated with the large shift in centromeric RT. In mammals, RT is considered a functional readout of large-scale chromatin structure [14, 26, 82], and regions that shift RT have been shown to also change 3D localization [118]. Likewise, a study in mouse showed that when late replicating pericentric heterochromatin was experimentally repositioned to the nuclear periphery, a location where mid replicating chromatin is usually found in that system, the RT of those regions was advanced [119].

Additionally, we speculate that centromere transcription could play a role in the shift to later replication of centromeres in the endocycle. Non-coding transcription has been reported

to occur in centromeres across a wide range of eukaryotes, and has also been found to be essential for centromere function ([112] and references therein). Furthermore, in human cells, when centromere RNAs transcribed from alpha satellite repeats are specifically degraded, cells arrest before mitosis and have reduced CENP-A levels [120]. In maize, low level transcription has been detected from *CentC* satellite repeats as well as *CRM* retroelements, and there is evidence that centromere transcripts are involved in the function of the kinetochore complex and centromere chromatin organization [121–123]. As noted above, technical limitations prevented us from assessing transcripts from endocycling cells. However, it seems possible that a reduction in centromere transcription during the endocycle could contribute to the observed reduction in CENH3 and potential dismantling of the kinetochore function. Outside the centromere region, reduced transcription is often associated with late replication [19, 26]. Future work addressing the technical challenge of sorting maize nuclei for transcript profiling will be needed to test this hypothesis.

One further interesting possibility arises from the observation that small regions in or adjacent to the 7 complex centromeres exhibit early replication during the mitotic cycle which is reduced in the endocycle. In mitotic cells, regions flanking these early peaks often show strong mid-S replication with relatively little replication in late S. In endocycling cells there is still very little late replication at the actual locus of the early peak, but late replication is enhanced in the flanking centromere regions. These small, early-replicating regions are reminiscent of the early replicating origins described in and near centromeres in several types of yeast [89, 101] as well as the "initiation regions" identified in very early S phase recently described in *Arabidopsis* by Wheeler et al. [105]. Strikingly, some of the *Arabidopsis* very early initiation regions occur in centromeric/pericentromeric heterochromatin of *Arabidopsis* chromosomes. It is possible to imagine that these early replicating regions at maize centromeres contain early-firing origins that drive replication of bulk centromeric chromatin during mid-S in mitotic cells, but that in endocycling cells, initiation activity in these regions is reduced. Alterations in chromatin conformation, subnuclear positioning, or other factors could mediate such a reduction in early initiation activity, which in turn could cause, or contribute to, the reduction in overall centromere transcription hypothesized in the previous paragraph.

The three speculative hypotheses presented above are not mutually exclusive, and in fact can easily be imagined to work synergistically together. We believe all three deserve further investigation as appropriate techniques can be developed. Investigating the interplay of chromatin environment, subnuclear organization, transcription and DNA replication in plant systems has proven difficult in the past. Numerous reasons for these difficulties exist, for example, plants have cell walls and are rich in nucleases, actively dividing cells are sequestered in tiny meristematic regions, and many genomes have a high content of retrotransposons and other repeats. As a result, understanding of such critical areas has lagged behind that in yeast and animal systems. However, with recent progress in assembling genomic resources and anticipated advances in the ability to isolate individual cell types [124], perform sophisticated analyses of genome conformation [125, 126] and follow individual chromosome regions using elegant cytological paints [107], the maize root tip system is poised to contribute to rapid progress in these and many other important areas of plant genome biology.

Methods

Plant material

Seeds of *Zea mays* inbred line B73 (GRIN NPGS PI 550473) were germinated on damp paper towels and grown for three days. Seedling roots were labeled by immersion in sterile water containing 25 μM EdU (Life Technologies) for 20 min, using growth and experimental

conditions described previously [8, 19, 20]. Biological replicate material was grown independently and harvested on different days. For the endocycle Repli-seq experiment, after rinsing roots well with sterile water, the 1–3 mm segments (Fig 1A) were excised from primary and seminal roots. The root segments were fixed, washed and snap-frozen as described previously [20].

Flow cytometry and sorting of root nuclei

Details of the flow sorting for Repli-seq analysis were described previously [19, 20]. Briefly, nuclei were isolated from the fixed root segments, and the incorporated EdU was conjugated to AF-488 using a Click-iT EdU Alexa Fluor 488 Imaging Kit (Life Technologies). The nuclei were then resuspended in cell lysis buffer (CLB) [20] containing 2 μg/mL DAPI and 40 μg/mL Ribonuclease A and filtered through a CellTrics 20-μm nylon mesh filter (Partec) just before flow sorting on an InFlux flow cytometer (BD Biosciences) equipped with UV (355 nm) and blue (488 nm) lasers. Nuclei prepared from the 1–3 mm root segments were sorted to collect populations of EdU/AF-488-labeled nuclei with DNA contents in three defined sub-stage gates between 4C and 8C, corresponding to early, mid and late S phase of the endocycle. The early endocycle gate was shifted slightly to the right to exclude mitotic nuclei in late S phase (Fig 1C). For each biological replicate, between 50,000 and 200,000 nuclei were sorted from each fraction of the endocycle S phase. A small sample of nuclei from each gate was sorted into CLB buffer containing DAPI and reanalyzed to determine the sort purity (S1 Fig). Sorting and reanalysis details for the mitotic nuclei are described in [19].

For ChIP-seq experiments, roots were labeled with EdU, and nuclei were isolated from 0–3 mm (H3K27me3 and H3K4me3) or 0–5 mm (H3K56ac) root segments and conjugated to AF-488 as described above. The 2C, 4C and 8C unlabeled, non S-phase populations of nuclei were sorted into 2× extraction buffer 2 (EB2) [127] using the same sorting conditions as in Wear et al. [19]. After sorting, the 2× EB2 was diluted to 1× with 1× STE. All flow cytometry data were analyzed using FlowJo v10.0.6 (TreeStar, Inc.) as described in Wear et al. [19].

DNA and chromatin immunoprecipitations

For endocycle Repli-seq samples, reversal of formaldehyde cross links, nuclear DNA purification and isolation, DNA shearing, EdU/AF-488 DNA immunoprecipitation with an anti-Alexa Fluor 488 antibody (Molecular Probes, #A-11094, lot 895897), and DNA fragment purification were performed as described in Wear et al. [19].

ChIP procedures were performed as in Wear et al. [19] except the chromatin was sheared using a Covaris S220 ultrasonicator to an average fragment size of 200 bp using a peak incident power of 140 W, 10% duty cycle, and 200 cycles per burst for 6 min. Three percent of the chromatin volume was set aside to use as the input control for each of the 2C, 4C and 8C samples and frozen at -70°C until the formaldehyde cross link reversal step. The antibodies used for ChIP were as follows: *Zea mays* anti-CENH3 antibody at a 1:250 dilution (gift from R.K. Dawe) [58], anti-H3K9me2 antibody at a 1:25 dilution (Cell Signaling Technologies; 9753, lot 4), anti-H3K56ac antibody at a 1:200 dilution (Millipore; 07–677, lot DAM1462569), anti-H3K4me3 antibody at a 1:300 dilution (Millipore; 07–473, lot DAM1779237) and anti-H3K27me3 antibody at a 1:300 dilution (Millipore; 07–449, lot 2,275,589). See S18 Fig for antibody validation experiments for anti-H3K9me2 and anti-CENH3.

Library construction and sequencing

For Repli-seq and ChIP-seq samples, the final purified DNA was used to construct paired-end libraries as described [19]. After adapter ligation, all samples underwent 17 cycles of PCR. For

each Repli-seq or ChIP-seq experiment, individual samples from three biological replicates collected on different days were barcoded, pooled and sequenced on either the Illumina HiSeq 2000 or NextSeq platforms. However, in the case of the Repli-seq mitotic late-S samples and CENH3 ChIP 4C samples, one biological replicate failed during library generation or sequencing, resulting in data from only two biological replicates. Repli-seq and ChIP-seq read mapping statistics are shown in S1 Spreadsheet.

Replication timing data analysis

Trimming and quality control of 100-bp paired-end Repli-seq reads were carried out as described previously [19], and reads were aligned to the maize B73 RefGen_v4 reference genome [46] (Ensembl Plants release 33; ftp://ftp.ensemblgenomes.org/pub/plants/release-33/gff3/zea_mays/) using BWA-MEM v0.7.12 with default parameters [128]. Redundant reads resulting from PCR amplification were removed from each of the alignment files using Picard (http://broadinstitute.github.io/picard/) and SAMtools [129]. Properly paired, uniquely mapping reads (MAPQ score > 10) were retained with SAMtools [129] for downstream analysis. The resulting mitotic Repli-seq data were more than 3× the sequencing coverage of the endocycle Repli-seq data (S1 Spreadsheet). Repli-seq results are robust at various sequencing depths [24], but to ensure that the mitotic and endocycle data were comparable, the reads were downsampled by a uniform random process using a custom python script incorporating the BEDTools suite [130] to a total of 65.7 million reads per S-phase fraction (S1 Spreadsheet). We preferred this to normalization so that any possible sampling bias due to sequencing depth would be similar in all samples.

Repli-seq data were analyzed using *Repliscan* [24]. Individual biological replicates of Repli-seq data were independently analyzed, and after finding good correlation between replicates (Pearson correlation coefficients from 0.80–0.99; S4 Fig) the replicates were aggregated by sum and normalized to 1× genome coverage using the reads per genomic content (RPGC) method. The following changes from the *Repliscan* default parameters described in [19] were used. Read densities were aggregated in 3-kb windows across the genome (parameter -*w* 3000). Additionally, we customized the cutoff for reducing type one errors which excluded genomic windows with extremely low coverage in the 2C reference sample. To identify these low read mapping windows, which we labeled "blacklist", *Repliscan* natural log-transformed the read counts from the pre-replicative 2C reference sample and windows with read counts in the lower 2.5% tail of a fitted normal distribution were excluded from all samples (parameter—*pcut* 2.5–100). The upper 2.5% tail containing extremely high coverage windows or "spikes" was not removed at this step, because we found that these data spikes were adequately normalized in the subsequent step of dividing each 3-kb window in the S-phase samples by the 2C reference data–which also normalized for sequencing biases and collapsed repeats (S3 Fig). The data were then Haar wavelet smoothed [24] to produce the final profiles for early, mid and late S-phase replication signals in the mitotic cycle and endocycle. Processed data files, formatted for the Integrative Genomics Viewer (IGV) [131], are available for download from CyVerse (formerly the iPlant Collaborative; [132]) via the information in S1 Spreadsheet.

Identifying regions of altered replication timing

The difference between normalized replication signal profiles of mitotic and endocycle Repli-seq data for early, mid, and late S was calculated in 3-kb windows, and the maximum negative and positive differences were then calculated for each chromosome and averaged. Regions showing a DRT of ≥ 25% (absolute difference in replication signal ≥ 1.0) or ≥ 10% (absolute difference in replication signal ≥ 0.4) of the total range of differences in each profile were

identified (S1 Table; S5 Fig) using the data filter tool in SAS JMP Pro v14 (SAS Institute Inc.). Windows were kept in the analysis only if their DRT were "compensated" by opposite DRT of ≥ 25% or ≥ 10%, respectively, in one or both of the other two S-phase fractions. For example, a decrease in early replication signal in endocycling cells must be compensated by an increase in mid and/or late S-phase signal in the same cell population. Adjacent 3-kb windows with DRT that met either the ≥ 10% or ≥ 25% threshold were merged, keeping the two files separate, using mergeBED in the BEDTools suite, and allowing a 6 kb gap distance (parameter -*d* 6000) [130]. This initial step resulted in many very small regions being identified (S2 Table). As a second step, if ≥ 10% regions were immediately adjacent to ≥ 25% regions, they were merged together using mergeBED to highlight larger regions of contiguous change (S2 Table). Only regions that contained at least one ≥ 25% region were kept for further analysis, and termed regions of alternate timing (RATs). By requiring a ≥ 25% DRT core region to be included, all of the stand-alone, extremely small regions (< 24 kb) were effectively filtered out, without the requirement of an arbitrary size filter. RATs were categorized into three groups: 1) later in mitotic to earlier in endocycle (Later-to-Earlier), 2) earlier in mitotic to later in endocycle (Earlier-to-Later) and 3) a subset of the Earlier-to-Later RATS that were located in the previously identified functional centromeres (Earlier-to-Later-CEN) (coordinates from [46]). There were no Later-to-Earlier-CEN RATs. For a list of RAT regions, including genomic coordinates and genes within them, see S2 and S3 Spreadsheets.

ChIP-seq data analysis

ChIP-seq reads for H3K27me3, H3K4me3, H3K56ac (100-bp paired-end reads), H3K9me2 and CENH3 (150-bp paired-end reads) were trimmed, mapped to maize B73 RefGen_v4.33, and filtered to retain only properly-paired, uniquely-mapped reads (MAPQ score > 10) as described above for Repli-seq reads. The 2C ChIP and input data for H3K27me3, H3K4me3, H3K56ac is from [19], while the 4C and 8C ChIP data was generated for this study, see S1 Spreadsheet. For details on peak calling and analysis for H3K27me3, H3K4me3, H3K56ac, see S1 Text.

For visualization of CENH3 localization in 2C, 4C and 8C nuclei, read counts for individual biological replicates of CENH3 or input samples were scaled to 1× genome coverage using the reads per genomic content (RPGC) method. Biological replicate data had good agreement (Pearson's correlation coefficient values between biological replicates of 0.97–0.99; S1 Spreadsheet), and were merged and scaled again to 1× coverage so the samples would be comparable. CENH3 scaled read counts in each 3-kb window were divided by the scaled read counts from the input sample for the corresponding ploidy level, resulting in CENH3 fold enrichment values relative to input.

To compare CENH3 enrichment relative to DNA content in 2C, 4C and 8C cells over entire centromeres, we calculated the percent of total CENH3 reads found in a given centromere (using coordinates from [46]), and divided by the percent of total input reads corresponding to that centromere. This was done separately for individual biological replicates; we then calculated the mean fold enrichment estimates. H3K9me2 fold enrichment over entire centromeres and pericentromeres was calculated in the same way.

Genomic features

The maize filtered gene set Zm00001d.2 annotation from B73 RefGen_v4 [46] was downloaded from Ensembl Plants (ftp://ftp.ensemblgenomes.org/pub/plants/release-33/gff3/zea_mays/). The updated B73 Refgen_v4 TEv2 disjoined annotation [47] was downloaded from http://mcstitzer.github.io/maize_TEs. Coordinates for mapped *CentC* satellite repeat regions

are described in Gent et al. [57]. The percent AT content was calculated in 3-kb static windows across the genome.

Analysis of features in RATs and random permutation analysis

We tested the association of various genomic features with the non-CEN RAT categories by determining the overlap of a particular feature with each RAT type. The coordinates for genomic features (genes, expressed genes, TE superfamilies) were intersected with RAT coordinate intervals using intersectBED (parameters -*wa* -*wb*) in the BEDtools suite [130]. The percent of RATs containing a feature or the percent coverage of TE superfamilies were then computed. The number of genes per RAT was also determined using intersectBED (parameter -*u*).

For comparison, the coordinates for the non-CEN Earlier-to-Later and Later-to-Earlier RAT sets were randomly shuffled around the genome, excluding functional centromeres, using BEDTools shuffle [130]. These random sets preserved the number of regions and region size of the original RAT sets, and are labeled "EtoL shuffle1" and "LtoE shuffle1" for the Earlier-to-Later and Later-to-Earlier RATs, respectively (e.g. S7, S8 and S10 Figs). When there appeared to be differences in the observed overlap values with genomic features between non-CEN RATs and their corresponding random shuffle sets, a permutation or feature randomization test, as described in [19] was used to assess the statistical significance of the observed value. To do so, the coordinates for the non-CEN RAT sets were randomly shuffled around the genome 1000 times, as described above, to generate an "expected" distribution. Permutation P values, calculated as described in [19], that were ≤ 0.001 were considered evidence that the observed percent coverage value is significantly different than random expectation.

Analysis of features in centromeres and pericentromeres

For comparison to CEN regions (coordinates from [46]), pericentromeres were arbitrarily defined as the ± 1 Mb flanking each CEN. In the case of chromosome 9, the pericentromere included the ± 1 Mb flanking both CEN 9a and 9b. Replication signal values in CENs and pericentromeres were intersected with genes, *CRM1* and *CRM2* families and mapped *CentC* regions using intersectBED (parameters -*wa* -*wb*) in the BEDtools suite [130]. Only elements that covered at least half of a 3-kb window of Repli-seq data were included in Fig 5, while elements with any amount of overlap were included in S14 Fig. Additionally, if a single gene or *CRM* element spanned more than one of the 3-kb windows, the replication signals were averaged using mergeBED (parameter -*o* mean) to compute a single value for the entire gene or element.

Supporting information

S1 Text. Supplemental Methods.
(DOCX)

S1 Fig. (related to Fig 1) Assessment of purity of flow sorted endocycling nuclei. Maize root tip nuclei were isolated from the 1–3 mm root region and sorted on a BD InFlux flow sorter. A small sample from each of the three S-phase sort gates was re-analyzed to determine the purity of the sorted nuclei. Histograms of relative DNA content (DAPI fluorescence) from re-analyzed sorted nuclei are overlaid for early (E), mid (M), and late (L) S-phase gates from the endocycle arc to show the separation between sorted samples. Similar separation was found for sorted early, mid and late nuclei from the mitotic cycle (see S1 Fig in [19]). The histogram of relative DNA content for the entire unsorted nuclei population (black line) is shown for reference.
(PDF)

S2 Fig. (related to Fig 1) Genomic copy number analysis. Whole genome sequence data from sorted non S-phase 2C, 4C and 8C nuclei were used to assess copy number per DNA content across the genome. To better represent the copy number of repeat regions, the primary alignment location for each read pair–even those that map to multiple locations–were included in the analysis. (**A** and **B**) Histograms of the normalized read frequency ratios, calculated in 5-kb static windows, for 2C/4C (**A**) and 8C/4C (**B**) nuclei. The black dashed lines indicate the overall mean and the red dashed lines indicate ± 2 S. D. from the mean. (**C**) The 8C/4C read frequency ratios plotted as a function of genomic location, which shows that the values outside ± 2 S. D. all occur as singleton 5-kb windows. (**D** and **E**) We used consensus sequences for 45S rDNA and *knob180* (**D**), and for 5S rDNA, *TR-1*, *CentC* and *CRM1–4* families (**E**) to individually query all of the trimmed whole genome sequence reads using BLAST software and a non-stringent E value to allow for variants of each repeat (S1 Text). The mean percentage of total reads that align to each repeat type was calculated for three biological replicates of 2C, 4C and 8C data. Black dots represent the individual biological replicate values. The apparent slight under-replication of several elements (e.g. *knob180* and *CRM2*) is not statistically significant.
(PDF)

S3 Fig. (related to Figs 1 and 4) Example of Repli-seq data processing with *Repliscan*. An example region from CEN 10 is shown to illustrate that the pre-replicative 2C reference data effectively normalizes spikes of signal in the S-phase data. (**A and B**) Read densities were calculated in 3-kb windows for the 2C reference (**A**) and each S-phase sample (endocycle late profile shown; **B**). After excluding blacklist regions (e.g. unmappable and multi-mapping regions), reads were scaled for overall sequence depth in each sample. (**C**) Scaled reads in each S-phase sample were normalized by making a ratio to 2C reference scaled reads in each 3-kb window. (**D**) Replication signal profiles were smoothed using a Haar wavelet transform to remove noise without altering peak boundaries.
(PDF)

S4 Fig. (related to Fig 1) Pearson's correlation coefficient values between individual biological replicates of mitotic and endocycle Repli-seq data. (**A and B**) Biological replicates (BR) of early (E), mid (M) and late (L) Repli-seq data for the mitotic cycle (Mit; panel **A**) and endocycle (En; panel **B**) was analyzed independently using *Repliscan* [24]. The agreement between biological replicates was assessed by calculating Pearson's correlation coefficients. (**C**) The Pearson's correlation coefficients for E, M, L data between mitotic cycle and endocycle.
(PDF)

S5 Fig. (related to Fig 2) Boxplots of differences in early, mid and late replication signal profiles for each chromosome. Differences in replication timing (DRT) signal were calculated by subtracting the mitotic signal from the endocycle signal for early (E), mid (M) and late (L) S-phase fractions in each 3-kb window across the genome. The distributions of DRT signal values are represented as violin plots for each chromosome. Median values are indicated by colored squares and 1.5 x IQR of the distribution is indicated by colored whisker lines. Dashed lines indicate the thresholds used in subsequent steps for identifying RATs ($\geq$ 10% and $\geq$ 25% of the total difference range; S1 Table).
(PDF)

S6 Fig. (related to Fig 2) Additional examples of non-CEN RATs. (**A–F**) Example regions on chromosomes 1 (**A**), 3 (**B**), 4 (**C**), 5 (**D**), 6 (**E**) and 7 (**F**) that include RATs. See main text Fig 2 legend for description. Dashed boxes denote regions with some level of DRT in which the magnitude of the difference did not meet our $\geq$ 25% criterion (boxes labeled "a" in panels **A**, **B**, **C** and **F**), or in which the change in one S-phase fraction was not compensated by an

opposite change in at least one other S-phase fraction (boxes labeled "b" in panels **C** and **D**). (PDF)

S7 Fig. (related to Fig 2) Activating and repressive histone marks in non-CEN RATs. To assess whether changes in selected histone modifications related to gene transcription and chromatin accessibility occur in RATs, ChIP-seq data was generated for H3K56ac and H3K4me3 (active transcription and early replication) and H3K27me (repressive transcription and facultative heterochromatin) from sorted non S-phase 2C, 4C and 8C nuclei. (**A–C**) The distributions of fold enrichment values for H3K56ac (**A**), H3K4me3 (**B**) and H3K27me3 (**C**) peaks in expressed and non-expressed genes (see S1 Text) in 2C, 4C and 8C nuclei are plotted as boxplots for Later-to-Earlier and Earlier-to-Later RATs and their corresponding randomly shuffled sets (see Methods). Asterisks indicate statistically significant differences by the non-parametric Steel-Dwass-Critchlow-Fligner test at the following P value levels: ***, $P < 0.0001$; **, $P < 0.001$; *, $P < 0.01$. The increase in the fold enrichment of H3K56ac for expressed genes in Earlier-to-Later RATs (panel **A**) may be associated with increases in peak enrichment we observed near the 3' end of some genes. (**D**) The count and percentage of expressed and non-expressed genes with each histone modification shown in the boxplots in panels **A–C**. The 8C/2C ratio of genes with each mark is also shown to demonstrate there is very little change in the number of genes with each mark. The total number of expressed and non-expressed genes in each RAT or random category are shown at the bottom for reference. (PDF)

S8 Fig. (related to Fig 2) Gene ontology analysis of genes in non-CEN RATs. Using the Plant GO slim ontology subset, we identified 44 significant GO terms in the biological process (P), molecular function (F), and cellular component (C) GO categories that were enriched in expressed genes (S1 Text; S3 Spreadsheet) in Earlier-to-Later RATs. Genes in the corresponding randomly shuffled set shared a few of the significantly enriched cellular component terms as genes in Earlier-to-Later RATs, suggesting that these terms may be related to common components of the root, and not RATs specifically. The total number of expressed genes in each input gene list was as follows: Later-to-Earlier RATs, 52; LtoE shuffle1 random regions, 68; Earlier-to-Later RATs, 292; EtoL shuffle1 random regions, 275. (PDF)

S9 Fig. (related to Fig 3) Permutation analysis of the percent coverage of TE superfamilies in non-CEN RATs. The percent coverage of all TEs and individual TE superfamilies annotated in the B73 RefGen_v4 genome was calculated for Earlier-to-Later (**A** and **B**) and Later-to-Earlier (**C** and **D**) non-CEN RATs and corresponding 1000 randomly shuffled sets (see Methods). The observed percentage for RATs (red or blue lines) are plotted alongside the expected frequency distribution of the random sets (grey violin plots overlaid with boxplots). Permutation P values below the graphs were calculated from the proportion of the 1000 random sets that have percent coverage values greater than (up arrow) or less than (down arrow) the observed value. No overlap was observed between Later-to-Earlier RATs and the RIT/LINE RTE superfamily, thus no P value was calculated. (PDF)

S10 Fig. (related to Fig 2) AT content composition in non-CEN RATs. (**A**) The distributions of percent AT content, calculated in 3-kb static windows, for Later-to-Earlier and Earlier-to-Later non-CEN RATs and the corresponding random shuffle sets are plotted as boxplots. Values outside the boxplot whiskers (1.5 x IQR) are represented as grey dots. The dashed line indicates the genome wide median value. (PDF)

S11 Fig. (related to Fig 4) Uniquely mapping Repli-seq reads in centromeres. The average percentage of centromeric reads that map to unique locations is shown for each Repli-seq sample. Black dots represent the individual values for biological replicates.
(PDF)

S12 Fig. (related to Fig 4) Replication signal profiles and RATs in complex and simple centromeres. 5-Mb regions are shown for complex CENs 2, 3, 4, 5, and 8 and simple CEN 7. See main text Fig 4 legend for description.
(PDF)

S13 Fig. (related to Fig 4) RT differences in centromeres and pericentromeres. DRT values (endocycle minus mitotic) were calculated from early (**A** and **D**), mid (**B** and **E**) and late (**C** and **F**) RT profiles for each centromere and corresponding pericentromere (± 1 Mb) in 100-kb static windows. In panels **D**, **E**, and **F** asterisks indicate DRT values from windows where an Earlier-to-Later-CEN RAT extends past the called CEN boundary [46] into the pericentromere; open circles indicate windows that contain a non-CEN Earlier-to-Later RAT that met our compensation criteria.
(PDF)

S14 Fig. (related to Fig 5) Replication times for all genomic features in complex and simple centromeres and corresponding pericentromeres. (**A–E**) The distributions of replication signals in early (E), mid (M), and late (L) during mitotic and endocycle S phases for all 3-kb windows (**A**), annotated genes (**B**), *CRM1* elements (**C**), *CRM2* elements (**D**), and mapped *CentC* repeats (**E**) in centromeres and pericentromeres (± 1 Mb). All elements within centromeres and pericentromeres are included, not just those that cover at least half of a 3-kb window, as in Fig 5. See main text Fig 5 legend for further description.
(PDF)

S15 Fig. (related to Figs 4 and 5) Activating and repressive histone mark peaks of enrichment in centromeres. ChIP-seq data were generated for H3K56ac, H3K4me3 (active transcription) and H3K27me (repressive transcription) from 2C, 4C and 8C nuclei. (**A–C**) The fold enrichment values for peaks in expressed and non-expressed genes for H3K56ac (**A**), H3K4me3 (**B**) and H3K27me3 (**C**) in 2C, 4C and 8C nuclei. Red lines indicate the median value. (**D**) The number of expressed and non-expressed genes with each mark in 2C, 4C and 8C nuclei.
(PDF)

S16 Fig. (related to Figs 4–6) H3K9me2 fold enrichment relative to DNA content in complex and simple centromeres. We used the ChIP-seq datasets from 2C, 4C and 8C nuclei to estimate the H3K9me2 average fold enrichment relative to DNA content by calculating the percent of total H3K9me2 reads found in a given centromere (**A** and **B**) using coordinates from [46] or pericentromere (**C** and **D**) and dividing by the percent of total input reads corresponding to that centromere or pericentromere. Black dots represent the individual values from biological replicates.
(PDF)

S17 Fig. (related to Fig 6) CENH3 localization and enrichment in mitotic and endocycling centromeres. (**A**) CENH3 localization patterns for 2C, 4C and 8C nuclei for CEN 1–CEN 8. (**B**) CENH3 average fold enrichment relative to DNA content for complex and simple centromeres. See main text Fig 6 for CEN 9 and CEN 10 localization patterns and legend description.
(PDF)

S18 Fig. (related to Fig 6) ChIP-qPCR antibody validations for anti-CENH3 and anti-H3K9me2 antibodies. The percentage of input (% IP) was calculated for various antibody dilutions and primer sets for the *Zea mays* anti-CENH3 antibody (**A**) and anti-H3K9me2 antibody (**B**). Black dots in panel **A** represent the individual values from two biological replicates. Positive control primer sets (*CRM2* and Copia retrotransposons) and negative control primer sets (18S rDNA and Actin1 UTR) were used. The no antibody control (NoAB) values are too small to see on the graph. See S1 Text for Supplemental Methods.
(PDF)

S1 Table. (related to Fig 2) Replication timing signal differences and thresholds. The difference in replication signal between mitotic and endocycle profiles (endocycle minus mitotic) was calculated for each 3-kb window across the genome. The maximum negative difference value, which indicates a higher signal in the mitotic cycle, and the maximum positive difference value, which indicates a higher signal in the endocycle, are shown for early and late profiles. The average total difference range between these two values was used to calculate percentage thresholds for identifying RATs (see S2 Table and main text).
(DOCX)

S2 Table. (related to Fig 2) Summary statistics of preliminary RAT calling steps. The thresholds from S1 Table ($\geq$ 10% or $\geq$ 25%) were used to identify regions with DRT in early or late S phase that were compensated by difference(s) with an opposite sign in one or both of the other two S-phase fractions (early + mid or mid + late) with greater than or equal to the same magnitude. The count, minimum, maximum and median region size, and the total coverage of the B73 RefGen_v4 genome are shown. Final robust RATs included at least one core region with a $\geq$ 25% DRT, but immediately adjacent regions of $\geq$ 10% DRT were merged together with the $\geq$ 25% regions to identify larger regions of contiguous change.
(DOCX)

S3 Table. (related to Figs 2 and 3) Gene summary in non-CEN RATs. The percent of RATs that contain genes, the total number of genes and expressed genes and the mean gene count per RAT are shown.
(DOCX)

S4 Table. (related to Figs 4 and 5) Compensated differences in RT in complex centromeres and corresponding pericentromeres. We calculated the total number of 3-kb windows in complex centromeres and pericentromeres ($\pm$ 1 Mb), as well as the number of windows that show DRT values that are compensated (threshold $\geq$ 10%) by equal and opposite shifts in the other two S-phase fractions.
(DOCX)

S5 Table. (related to Fig 6) CENH3 average fold enrichment relative to DNA content in centromeres. CENH3 fold enrichment relative to DNA content and the ratio of enrichments between 4C and 2C and 8C and 2C are shown for each centromere. Fold enrichment values are the mean $\pm$ S. D. of three biological replicates for 2C and 8C and two biological replicates of 4C. See main text Fig 6 legend for further description. Two sets of theoretical ratio values are also presented. The first set, labeled "proportional redeposition", corresponds to the hypothesis that CENH3 is diluted relative to total DNA during replication, and is then redeposited to a level proportional to the DNA content during the subsequent gap phase. The second set, labeled "no redeposition", corresponds to an alternate hypothesis that CENH3 is diluted relative to total DNA during replication, and is not redeposited in the subsequent gap phase.
(DOCX)

S1 Spreadsheet. (related to Figs 1–6) Mapping statistics and data availability for all included datasets.
(XLSX)

S2 Spreadsheet. (related to Figs 2 and 4) RAT regions list.
(XLSX)

S3 Spreadsheet. (related to Figs 2–4) Genes found in RATs.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Emily E. Wear, Linda Hanley-Bowdoin, William F. Thompson.

Data curation: Emily E. Wear, Jawon Song, Leigh Mickelson-Young.

Formal analysis: Emily E. Wear, Jawon Song, Gregory J. Zynda.

Funding acquisition: George C. Allen, Robert A. Martienssen, Matthew W. Vaughn, Linda Hanley-Bowdoin, William F. Thompson.

Investigation: Emily E. Wear, Leigh Mickelson-Young, Chantal LeBlanc, Tae-Jin Lee, David O. Deppong.

Methodology: Emily E. Wear, Jawon Song, Gregory J. Zynda, Leigh Mickelson-Young, Tae-Jin Lee, Linda Hanley-Bowdoin, William F. Thompson.

Project administration: Linda Hanley-Bowdoin, William F. Thompson.

Resources: George C. Allen, Robert A. Martienssen, Matthew W. Vaughn, Linda Hanley-Bowdoin, William F. Thompson.

Software: Jawon Song, Gregory J. Zynda.

Supervision: Linda Hanley-Bowdoin, William F. Thompson.

Validation: Emily E. Wear, Jawon Song, Gregory J. Zynda.

Visualization: Emily E. Wear.

Writing – original draft: Emily E. Wear, William F. Thompson.

Writing – review & editing: Emily E. Wear, Jawon Song, Gregory J. Zynda, Leigh Mickelson-Young, Linda Hanley-Bowdoin, William F. Thompson.

References

1. Edgar BA, Zielke N, Gutierrez C. Endocycles: A recurrent evolutionary innovation for post-mitotic cell growth. Nat Rev Mol Cell Biol. 2014; 15: 197–210. https://doi.org/10.1038/nrm3756 PMID: 24556841

2. Lee HO, Davidson JM, Duronio RJ. Endoreplication: Polyploidy with purpose. Genes Dev. 2009; 23: 2461–2477. https://doi.org/10.1101/gad.1829209 PMID: 19884253

3. Fox DT, Duronio RJ. Endoreplication and polyploidy: Insights into development and disease. Development. 2013; 140: 3–12. https://doi.org/10.1242/dev.080531 PMID: 23222436

4. Galbraith DW, Harkins KR, Knapp S. Systemic endopolyploidy in *Arabidopsis thaliana*. Plant Physiol. 1991; 96: 985–989. https://doi.org/10.1104/pp.96.3.985 PMID: 16668285

5. Joubes J, Chevalier C. Endoreduplication in higher plants. Plant Mol Biol. 2000; 43: 735–745. https://doi.org/10.1023/a:1006446417196 PMID: 11089873

6. Breuer C, Ishida T, Sugimoto K. Developmental control of endocycles and cell growth in plants. Curr Opin Plant Biol. 2010; 13: 654–660. https://doi.org/10.1016/j.pbi.2010.10.006 PMID: 21094078

7. Hayashi K, Hasegawa J, Matsunaga S. The boundary of the meristematic and elongation zones in roots: Endoreduplication precedes rapid cell expansion. Sci Rep. 2013; 3: 2723. https://doi.org/10.1038/srep02723 PMID: 24121463

8. Bass HW, Wear EE, Lee TJ, Hoffman GG, Gumber HK, Allen GC, et al. A maize root tip system to study DNA replication programmes in somatic and endocycling nuclei during plant development. J Exp Bot. 2014; 65: 2747–2756. https://doi.org/10.1093/jxb/ert470 PMID: 24449386

9. Baluska F. Nuclear size, DNA content, and chromatin condensation are different in individual tissues of the maize root apex. Protoplasma. 1990; 158: 45–52.

10. Alarcón MV, Salguero J. Transition zone cells reach G2 phase before initiating elongation in maize root apex. Biology Open. 2017; 6: 909–913. https://doi.org/10.1242/bio.025015 PMID: 28495964

11. Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, et al. Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol. 2008; 6: e245. https://doi.org/10.1371/journal.pbio.0060245 PMID: 18842067

12. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. P Natl Acad Sci USA. 2010; 107: 139–144.

13. Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, Papp B, et al. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. 2010; 20: 155–169. https://doi.org/10.1101/gr.099796.109 PMID: 19952138

14. Rivera-Mulia JC, Buckley Q, Sasaki T, Zimmerman J, Didier RA, Nazor K, et al. Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. Genome Res. 2015; 25: 1091–1103. https://doi.org/10.1101/gr.187989.114 PMID: 26055160

15. Birnbaum K, Jung JW, Wang JY, Lambert GM, Hirst JA, Galbraith DW, et al. Cell type–specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. Nat Methods. 2005; 2: 615–619. https://doi.org/10.1038/nmeth0805-615 PMID: 16170893

16. Iyer-Pascuzzi AS, Benfey PN. Fluorescence-activated cell sorting in plant developmental biology. In: Hennig L, Köhler C, editors. Plant Developmental Biology, Methods in Mol Biol. New York: Humana Press; 2010. pp. 313–319.

17. Tanurdzic M, Vaughn MW, Jiang H, Lee TJ, Slotkin RK, Sosinski B, et al. Epigenomic consequences of immortalized plant cell suspension culture. PLoS Biol. 2008; 6: 2880–2895. https://doi.org/10.1371/journal.pbio.0060302 PMID: 19071958

18. Sugiyama M. Historical review of research on plant cell dedifferentiation. J Plant Res. 2015; 128: 349–359. https://doi.org/10.1007/s10265-015-0706-y PMID: 25725626

19. Wear EE, Song J, Zynda GJ, LeBlanc C, Lee TJ, Mickelson-Young L, et al. Genomic analysis of the DNA replication timing program during mitotic S phase in maize (*Zea mays*) root tips. Plant Cell. 2017; 29: 2126–2149. https://doi.org/10.1105/tpc.17.00037 PMID: 28842533

20. Wear EE, Concia L, Brooks AM, Markham EA, Lee T-J, Allen GC, et al. Isolation of plant nuclei at defined cell cycle stages using EdU labeling and flow cytometry. In: Caillaud M-C, editor. Plant Cell Division: Methods and Protocols, Methods in Mol Biol. New York: Humana Press; 2016. pp. 69–86.

21. Concia L, Brooks AM, Wheeler E, Zynda GJ, Wear EE, LeBlanc C, et al. Genome-wide analysis of the *Arabidopsis* replication timing program. Plant Physiol. 2018; 176: 2166–2185. https://doi.org/10.1104/pp.17.01537 PMID: 29301956

22. Bass HW, Hoffman GG, Lee TJ, Wear EE, Joseph SR, Allen GC, et al. Defining multiple, distinct, and shared spatiotemporal patterns of DNA replication and endoreduplication from 3D image analysis of developing maize Z*ea mays* L.) root tip nuclei. Plant Mol Biol. 2015; 89: 339–351. https://doi.org/10.1007/s11103-015-0364-4 PMID: 26394866

23. Savadel SD, Bass HW. Take a look at plant DNA replication: Recent insights and new questions. Plant Signal Behav. 2017; 12: e1311437. https://doi.org/10.1080/15592324.2017.1311437 PMID: 28375043

24. Zynda GJ, Song J, Concia L, Wear EE, Hanley-Bowdoin L, Thompson WF, et al. Repliscan: A tool for classifying replication timing regions. BMC Bioinformatics. 2017; 18: 362. https://doi.org/10.1186/s12859-017-1774-x PMID: 28784090

25. Pryor A, Faulkner K, Rhoades MM, Peacock WJ. Asynchronous replication of heterochromatin in maize. P Natl Acad Sci USA. 1980; 77: 6705–6709.

26. Marchal C, Sima J, Gilbert DM. Control of DNA replication timing in the 3D genome. Nat Rev Mol Cell Biol. 2019; 20: 721–737. https://doi.org/10.1038/s41580-019-0162-y PMID: 31477886

27. Hua BL, Orr-Weaver TL. DNA replication control during *Drosophila* development: Insights into the onset of S phase, replication initiation, and fork progression. Genetics. 2017; 207: 29–47. https://doi.org/10.1534/genetics.115.186627 PMID: 28874453

28. Hammond MP, Laird CD. Chromosome structure and DNA replication in nurse and follicle cells of *Drosophila melanogaster*. Chromosoma. 1985; 91: 267–278. https://doi.org/10.1007/BF00328222 PMID: 3920017

29. Hammond MP, Laird CD. Control of DNA replication and spatial distribution of defined DNA sequences in salivary gland cells of *Drosophila melanogaster*. Chromosoma. 1985; 91: 279–286. https://doi.org/10.1007/BF00328223 PMID: 3920018

30. Nordman J, Li S, Eng T, Macalpine D, Orr-Weaver TL. Developmental control of the DNA replication and transcription programs. Genome Res. 2011; 21: 175–181. https://doi.org/10.1101/gr.114611.110 PMID: 21177957

31. Hribova E, Holusova K, Travnicek P, Petrovska B, Ponert J, Simkova H, et al. The enigma of progressively partial endoreplication: New insights provided by flow cytometry and next-generation sequencing. Genome Biol Evol. 2016; 8: 1996–2005. https://doi.org/10.1093/gbe/evw141 PMID: 27324917

32. Travnicek P, Certner M, Ponert J, Chumova Z, Jersakova J, Suda J. Diversity in genome size and GC content shows adaptive potential in orchids and is closely linked to partial endoreplication, plant life-history traits and climatic conditions. New Phytol. 2019; 224: 1642–1656. https://doi.org/10.1111/nph.15996 PMID: 31215648

33. Bauer MJ, Birchler JA. Organization of endoreduplicated chromosomes in the endosperm of *Zea mays* L. Chromosoma. 2006; 115: 383–394. https://doi.org/10.1007/s00412-006-0068-2 PMID: 16741707

34. Jacob Y, Stroud H, Leblanc C, Feng S, Zhuo L, Caro E, et al. Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature. 2010; 466: 987–991. https://doi.org/10.1038/nature09290 PMID: 20631708

35. Salic A, Mitchison TJ. A chemical method for fast and sensitive detection of DNA synthesis *in vivo*. P Natl Acad Sci USA. 2008; 105: 2415–2420.

36. Mickelson-Young L, Wear E, Mulvaney P, Lee TJ, Szymanski ES, Allen G, et al. A flow cytometric method for estimating S-phase duration in plants. J Exp Bot. 2016; 67: 6077–6087. https://doi.org/10.1093/jxb/erw367 PMID: 27697785

37. Yarosh W, Spradling AC. Incomplete replication generates somatic DNA alterations within *Drosophila* polytene salivary gland cells. Genes Dev. 2014; 28: 1840–1855. https://doi.org/10.1101/gad.245811.114 PMID: 25128500

38. Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ. Highly repeated DNA sequence limited to knob heterochromatin in maize. P Natl Acad Sci USA. 1981; 78: 4490–4494.

39. Ananiev EV, Phillips RL, Rines HW. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: Are chromosome knobs megatransposons? P Natl Acad Sci USA. 1998; 95: 10785–10790.

40. Rivin CJ, Cullis CA, Walbot V. Evaluating quantitative variation in the genome of *Zea mays*. Genetics. 1986; 113: 1009–1019. PMID: 3744025

41. Ananiev EV, Phillips RL, Rines HW. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. P Natl Acad Sci USA. 1998; 95: 13073–13078.

42. Presting GG, Malysheva L, Fuchs J, Schubert I. A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. 1998; 16: 721–728. https://doi.org/10.1046/j.1365-313x.1998.00341.x PMID: 10069078

43. Miller JT, Dong FG, Jackson SA, Song J, Jiang JM. Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics. 1998; 150: 1615–1623. PMID: 9832537

44. Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, et al. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet. 2009; 5: e1000732. https://doi.org/10.1371/journal.pgen.1000732 PMID: 19936065

45. Gent JI, Schneider KL, Topp CN, Rodriguez C, Presting GG, Dawe RK. Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning. Epigenet Chromatin. 2011; 4: 3.

46. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. Nature. 2017; 546: 524–527. https://doi.org/10.1038/nature22971 PMID: 28605751

47. Anderson SN, Stitzer MC, Brohammer AB, Zhou P, Noshay JM, O'Connor CH, et al. Transposable elements contribute to dynamic genome content in maize. Plant J. 2019; 100: 1052–1065. https://doi.org/10.1111/tpj.14489 PMID: 31381222

48. Zhao PA, Sasaki T, Gilbert DM. High-resolution Repli-seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. Genome Biol. 2020; 21: 1–20.

49. Van't Hof J. DNA replication in plants. In: DePamphilis M, editor. DNA replication in eukaryotic cells. New York: Cold Spring Harbor Laboratory Press; 1996. pp. 1005–1014.

50. SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome. Science. 1996; 274: 765–768. https://doi.org/10.1126/science.274.5288.765 PMID: 8864112

51. Liu R, Vitte C, Ma J, Mahama AA, Dhliwayo T, Lee M, et al. A GeneTrek analysis of the maize genome. P Natl Acad Sci USA. 2007; 104: 11844–11849.

52. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: Complexity, diversity, and dynamics. Science. 2009; 326: 1112–1115. https://doi.org/10.1126/science.1178534 PMID: 19965430

53. Comai L, Maheshwari S, Marimuthu MPA. Plant centromeres. Curr Opin Plant Biol. 2017; 36: 158–167. https://doi.org/10.1016/j.pbi.2017.03.003 PMID: 28411416

54. Fukagawa T, Earnshaw WC. The centromere: Chromatin foundation for the kinetochore machinery. Dev Cell. 2014; 30: 497–509.

55. McKinley KL, Cheeseman IM. The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol. 2016; 17: 16–29. https://doi.org/10.1038/nrm.2015.5 PMID: 26601620

56. Gent JI, Nannas NJ, Liu Y, Su H, Zhao H, Gao Z, et al. Genomics of maize centromeres. In: Bennetzen J, Flint-Garcia S, Hirsch C, Tuberosa R, editors. The Maize Genome. Compendium of Plant Genomes. Cham: Springer; 2018. p. 59–80.

57. Gent JI, Wang N, Dawe RK. Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. Genome Biol. 2017;18. https://doi.org/10.1186/s13059-017-1249-4 PMID: 28637491

58. Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, et al. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell. 2002; 14: 2825–2836. https://doi.org/10.1105/tpc.006106 PMID: 12417704

59. Presting GG. Centromeric retrotransposons and centromere function. Curr Opin Genet Dev. 2018; 49: 79–84. https://doi.org/10.1016/j.gde.2018.03.004 PMID: 29597064

60. Zhao H, Zhu X, Wang K, Gent JI, Zhang W, Dawe RK, et al. Gene expression and chromatin modifications associated with maize centromeres. G3-Genes Genom Genet. 2015; 6: 183–192.

61. Zhang W, Lee HR, Koo DH, Jiang J. Epigenetic modification of centromeric chromatin: Hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. Plant Cell. 2008; 20: 25–34. https://doi.org/10.1105/tpc.107.057083 PMID: 18239133

62. Gent JI, Madzima TF, Bader R, Kent MR, Zhang X, Stam M, et al. Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. Plant Cell. 2014; 26: 4903–4917. https://doi.org/10.1105/tpc.114.130427 PMID: 25465407

63. Maheshwari S, Tan EH, West A, Franklin FCH, Comai L, Chan SW. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet. 2015; 11: e1004970. https://doi.org/10.1371/journal.pgen.1004970 PMID: 25622028

64. Shelby RD, Monier K, Sullivan KF. Chromatin assembly at kinetochores is uncoupled from DNA replication. J Cell Biol. 2000; 151: 1113–1118. https://doi.org/10.1083/jcb.151.5.1113 PMID: 11086012

65. Jansen LE, Black BE, Foltz DR, Cleveland DW. Propagation of centromeric chromatin requires exit from mitosis. J Cell Biol. 2007; 176: 795–805. https://doi.org/10.1083/jcb.200701066 PMID: 17339380

66. Boyarchuk E, Montes de Oca R, Almouzni G. Cell cycle dynamics of histone variants at the centromere, a model for chromosomal landmarks. Curr Opin Cell Biol. 2011; 23: 266–276. https://doi.org/10.1016/j.ceb.2011.03.006 PMID: 21470840

67. Nagaki K, Kashihara K, Murata M. Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant *Luzula nivea*. Plant Cell. 2005; 17: 1886–1893. https://doi.org/10.1105/tpc.105.032961 PMID: 15937225

68. Lermontova I, Schubert V, Fuchs J, Klatte S, Macas J, Schubert I. Loading of *Arabidopsis* centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. Plant Cell. 2006; 18: 2443–2451. https://doi.org/10.1105/tpc.106.043174 PMID: 17028205

69. Lermontova I, Fuchs J, Schubert V, Schubert I. Loading time of the centromeric histone H3 variant differs between plants and animals. Chromosoma. 2007; 116: 507–510. https://doi.org/10.1007/s00412-007-0122-8 PMID: 17786463

70. Schubert V, Lermontova I, Schubert I. Loading of the centromeric histone H3 variant during meiosis–how does it differ from mitosis? Chromosoma. 2014; 123: 491–497. https://doi.org/10.1007/s00412-014-0466-9 PMID: 24806806

71. Nechemia-Arbely Y, Miga KH, Shoshani O, Aslanian A, McMahon MA, Lee AY, et al. DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres. Nat Cell Biol. 2019; 21: 743–754. https://doi.org/10.1038/s41556-019-0331-4 PMID: 31160708

72. Sugimoto-Shirasu K, Roberts K. "Big it up": Endoreduplication and cell-size control in plants. Curr Opin Plant Biol. 2003; 6: 544–553. https://doi.org/10.1016/j.pbi.2003.09.009 PMID: 14611952

73. Bhosale R, Boudolf V, Cuevas F, Lu R, Eekhout T, Hu Z, et al. A spatiotemporal DNA endoploidy map of the *Arabidopsis* root reveals roles for the endocycle in root development and stress adaptation. The Plant Cell. 2018; 30: 2330–2351. https://doi.org/10.1105/tpc.17.00983 PMID: 30115738

74. Cooper S. Rethinking synchronization of mammalian cells for cell cycle analysis. Cell Mol Life Sci. 2003; 60: 1099–1106. https://doi.org/10.1007/s00018-003-2253-2 PMID: 12861378

75. Phillips RL, Kaeppler SM, Olhoft P. Genetic instability of plant tissue cultures: Breakdown of normal controls. P Natl Acad Sci USA. 1994; 91: 5222–5226.

76. Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell. 2010; 7: 521–531. https://doi.org/10.1016/j.stem.2010.07.017 PMID: 20887957

77. Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, et al. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell. 2011; 8: 106–118. https://doi.org/10.1016/j.stem.2010.12.003 PMID: 21211785

78. Armstrong RL, Das S, Hill CA, Duronio RJ, Nordman JT. Rif1 functions in a tissue-specific manner to control replication timing through its PP1-binding motif. Genetics. 2020; 215: 75–87. https://doi.org/10.1534/genetics.120.303155 PMID: 32144132

79. Dimitrova DS, Gilbert DM. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. Molecular Cell. 1999; 4: 983–993. https://doi.org/10.1016/s1097-2765(00)80227-0 PMID: 10635323

80. Frawley LE, Orr-Weaver TL. Polyploidy. Curr Biol. 2015; 25: R353–R358. https://doi.org/10.1016/j.cub.2015.03.037 PMID: 25942544

81. Schwaiger M, Stadler MB, Bell O, Kohler H, Oakeley EJ, Schubeler D. Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. Genes Dev. 2009; 23: 589–601. https://doi.org/10.1101/gad.511809 PMID: 19270159

82. Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. 2010; 20: 761–770. https://doi.org/10.1101/gr.099655.109 PMID: 20430782

83. Lubelsky Y, Prinz JA, DeNapoli L, Li YL, Belsky JA, MacAlpine DM. DNA replication and transcription programs respond to the same chromatin cues. Genome Res. 2014; 24: 1102–1114. https://doi.org/10.1101/gr.160010.113 PMID: 24985913

84. Armstrong RL, Penke TJR, Strahl BD, Matera AG, McKay DJ, MacAlpine DM, et al. Chromatin conformation and transcriptional activity are permissive regulators of DNA replication initiation in *Drosophila*. Genome Res. 2018; 28: 1688–1700. https://doi.org/10.1101/gr.239913.118 PMID: 30279224

85. Massey DJ, Kim D, Brooks KE, Smolka MB, Koren A. Next-generation sequencing enables spatiotemporal resolution of human centromere replication timing. Genes. 2019; 10: 269.

86. McCarroll RM, Fangman WL. Time of replication of yeast centromeres and telomeres. Cell. 1988; 54: 505–513. https://doi.org/10.1016/0092-8674(88)90072-4 PMID: 3042152

87. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, et al. Replication dynamics of the yeast genome. Science. 2001; 294: 115–121. https://doi.org/10.1126/science.294.5540.115 PMID: 11588253

88. Kim SM, Dubey DD, Huberman JA. Early-replicating heterochromatin. Genes Dev. 2003; 17: 330–335. https://doi.org/10.1101/gad.1046203 PMID: 12569122

89. Koren A, Tsai HJ, Tirosh I, Burrack LS, Barkai N, Berman J. Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. PLoS Genet. 2010; 6: e1001068. https://doi.org/10.1371/journal.pgen.1001068 PMID: 20808889

90. Tenhagen KG, Gilbert DM, Willard HF, Cohen SN. Replication timing of DNA-sequences associated with human centromeres and telomeres. Mol Cell Biol. 1990; 10: 6348–6355. https://doi.org/10.1128/mcb.10.12.6348 PMID: 2247059

91. O'Keefe R, Henderson S, Spector DL. Dynamic organization of DNA replication in mammalian cell nuclei: Spatially and temporally defined replication of chromosome-specific satelite DNA sequences. J Cell Biol. 1992;116.

92. Hultdin M, Gronlund E, Norrback KF, Just T, Taneja K, Roos G. Replication timing of human telomeric DNA and other repetitive sequences analyzed by fluorescence in situ hybridization and flow cytometry. Exp Cell Res. 2001; 271: 223–229. https://doi.org/10.1006/excr.2001.5391 PMID: 11716534

93. Sullivan B, Karpen G. Centromere identity in *Drosophila* is not determined in vivo by replication timing. J Cell Biol. 2001; 154: 683–690. https://doi.org/10.1083/jcb.200103001 PMID: 11514585

94. Ouspenski II, Van Hooser AA, Brinkley BR. Relevance of histone acetylation and replication timing for deposition of centromeric histone CENP-A. Exp Cell Res. 2003; 285: 175–188. https://doi.org/10.1016/s0014-4827(03)00011-9 PMID: 12706113

95. Shang WH, Hori T, Martins NM, Toyoda A, Misu S, Monma N, et al. Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev Cell. 2013; 24: 635–648. https://doi.org/10.1016/j.devcel.2013.02.009 PMID: 23499358

96. Fuchs J, Strehl S, Brandes A, Schweizer D, Schubert I. Molecular—cytogenetic characterization of the *Vicia faba* genome—heterochromatin differentiation, replication patterns and sequence localization. Chromosome Res. 1998; 6: 219–230. https://doi.org/10.1023/a:1009215802737 PMID: 9609666

97. Schubert I. Late-replicating satellites: Something for all centromeres? Trends Genet. 1998; 14: 385–386. https://doi.org/10.1016/s0168-9525(98)01570-4 PMID: 9820025

98. Jasencakova Z, Meister A, Schubert I. Chromatin organization and its relation to replication and histone acetylation during the cell cycle in barley. Chromosoma. 2001; 110: 83–92. https://doi.org/10.1007/s004120100132 PMID: 11453558

99. Samaniego R, de la Torre C, de la Espina SMD. Dynamics of replication foci and nuclear matrix during S phase in *Allium cepa* L. cells. Planta. 2002; 215: 195–204. https://doi.org/10.1007/s00425-002-0741-5 PMID: 12029468

100. Müller CA, Nieduszynski CA. Conservation of replication timing reveals global and local regulation of replication origin activity. Genome Res. 2012; 22: 1953–1962. https://doi.org/10.1101/gr.139477.112 PMID: 22767388

101. Pohl TJ, Brewer BJ, Raghuraman MK. Functional centromeres determine the activation time of pericentric origins of DNA replication in *Saccharomyces cerevisiae.* PLoS Genet. 2012; 8: e1002677. https://doi.org/10.1371/journal.pgen.1002677 PMID: 22589733

102. Natsume T, Müller CA, Katou Y, Retkute R, Gierliński M, Araki H, et al. Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. Mol Cell. 2013; 50: 661–674. https://doi.org/10.1016/j.molcel.2013.05.011 PMID: 23746350

103. Macheret M, Halazonetis TD. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. Nature. 2018; 555: 112–116. https://doi.org/10.1038/nature25507 PMID: 29466339

104. Smith OK, Kim R, Fu HQ, Martin MM, Lin CM, Utani K, et al. Distinct epigenetic features of differentiation-regulated replication origins. Epigenet Chromatin. 2016; 9: 18.

105. Wheeler E, Brooks AM, Concia L, Vera D, Wear EE, LeBlanc C, et al. *Arabidopsis* DNA replication initiates in intergenic, AT-rich open chromatin. Plant Physiol. 2020; 183: 206–220. https://doi.org/10.1104/pp.19.01520 PMID: 32205451

106. Albert PS, Gao Z, Danilova TV, Birchler JA. Diversity of chromosomal karyotypes in maize and its relatives. Cytogenet Genome Res. 2010; 129: 6–16. https://doi.org/10.1159/000314342 PMID: 20551613

107. Albert PS, Zhang T, Semrau K, Rouillard J-M, Kao Y-H, Wang C-JR, et al. Whole-chromosome paints in maize reveal rearrangements, nuclear domains, and chromosomal relationships. P Natl Acad Sci USA. 2019; 116: 1679–1685.

108. Hemmerich P, Weidtkamp-Peters S, Hoischen C, Schmiedeberg L, Erliandri I, Diekmann S. Dynamics of inner kinetochore assembly and maintenance in living cells. J Cell Biol. 2008; 180: 1101–1114. https://doi.org/10.1083/jcb.200710052 PMID: 18347072

109. Dembinsky D, Woll K, Saleem M, Liu Y, Fu Y, Borsuk LA, et al. Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. Plant Physiol. 2007; 145: 575–588. https://doi.org/10.1104/pp.107.106203 PMID: 17766395

110. Orr-Weaver TL. When bigger is better: The role of polyploidy in organogenesis. Trends Genet. 2015; 31: 307–315. https://doi.org/10.1016/j.tig.2015.03.011 PMID: 25921783

111. del Arco AG, Edgar BA, Erhardt S. In vivo analysis of centromeric proteins reveals a stem cell-specific asymmetry and an essential role in differentiated, non-proliferating cells. Cell Rep. 2018; 22: 1982–1993. https://doi.org/10.1016/j.celrep.2018.01.079 PMID: 29466727

112. Wong CYY, Lee BCH, Yuen KWY. Epigenetic regulation of centromere function. Cell Mol Life Sci. 2020: 1–19.

113. Muller S, Almouzni G. Chromatin dynamics during the cell cycle at centromeres. Nat Rev Genet. 2017; 18: 192–208. https://doi.org/10.1038/nrg.2016.157 PMID: 28138144

114. Zhang XL, Li XX, Marshall JB, Zhong CX, Dawe RK. Phosphoserines on maize centromeric histone H3 and histone H3 demarcate the centromere and pericentromere during chromosome segregation. Plant Cell. 2005; 17: 572–583. https://doi.org/10.1105/tpc.104.028522 PMID: 15659628

115. Demidov D, Heckmann S, Weiss O, Rutten T, Tomastikova ED, Kuhlmann M, et al. Deregulated phosphorylation of CENH3 at Ser65 affects the development of floral meristems in *Arabidopsis thaliana*. Frontiers in Plant Science. 2019;10. https://doi.org/10.3389/fpls.2019.00928 PMID: 31404279

116. Niikura Y, Kitagawa R, Ogi H, Abdulle R, Pagala V, Kitagawa K. CENP-A K124 ubiquitylation is required for CENP-A deposition at the centromere. Dev Cell. 2015; 32: 589–603. https://doi.org/10.1016/j.devcel.2015.01.024 PMID: 25727006

117. Bui M, Pitman M, Nuccio A, Roque S, Donlin-Asp PG, Nita-Lazar A, et al. Internal modifications in the CENP-A nucleosome modulate centromeric dynamics. Epigenet Chromatin. 2017; 10: 17.

118. Takebayashi S, Dileep V, Ryba T, Dennis JH, Gilbert DM. Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. P Natl Acad Sci USA. 2012; 109: 12574–12579.

119. Heinz KS, Casas-Delucchi CS, Torok T, Cmarko D, Rapp A, Raska I, et al. Peripheral re-localization of constitutive heterochromatin advances its replication timing and impairs maintenance of silencing marks. Nucleic Acids Res. 2018; 46: 6112–6128. https://doi.org/10.1093/nar/gky368 PMID: 29750270

120. McNulty SM, Sullivan LL, Sullivan BA. Human centromeres produce chromosome-specific and array-specific alpha satellite transcripts that are complexed with CENP-A and CENP-C. Dev Cell. 2017; 42: 226–240. e226. https://doi.org/10.1016/j.devcel.2017.07.001 PMID: 28787590

121. Topp CN, Zhong CX, Dawe RK. Centromere-encoded RNAs are integral components of the maize kinetochore. P Natl Acad Sci USA. 2004; 101: 15986–15991.

122. Du Y, Topp CN, Dawe RK. DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. PLoS Genet. 2010; 6: e1000835. https://doi.org/10.1371/journal.pgen.1000835 PMID: 20140237

123. Liu Y, Su H, Zhang J, Liu Y, Feng C, Han F. Back-spliced RNA from retrotransposon binds to centromere and regulates centromeric chromatin loops in maize. PLoS biology. 2020; 18: e3000582. https://doi.org/10.1371/journal.pbio.3000582 PMID: 31995554

124. Krishnakumar V, Choi Y, Beck E, Wu Q, Luo A, Sylvester A, et al. A maize database resource that captures tissue-specific and subcellular-localized gene expression, via fluorescent tags and confocal imaging (Maize Cell Genomics Database). Plant Cell Physiol. 2015; 56: e12. https://doi.org/10.1093/pcp/pcu178 PMID: 25432973

125. Dong P, Tu X, Chu PY, Lu P, Zhu N, Grierson D, et al. 3d chromatin architecture of large plant genomes determined by local a/b compartments. Mol Plant. 2017; 10: 1497–1509. https://doi.org/10.1016/j.molp.2017.11.005 PMID: 29175436

126. Sotelo-Silveira M, Chavez Montes RA, Sotelo-Silveira JR, Marsch-Martinez N, de Folter S. Entering the next dimension: Plant genomes in 3d. Trends Plant Sci. 2018; 23: 598–612. https://doi.org/10.1016/j.tplants.2018.03.014 PMID: 29703667

127. Gendrel AV, Lippman Z, Martienssen R, Colot V. Profiling histone modification patterns in plants using genomic tiling microarrays. Nat Methods. 2005; 2: 213–218. https://doi.org/10.1038/nmeth0305-213 PMID: 16163802

128. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [Preprint]. arXiv:13033997. 2013 [cited 2020 July 1]. Available from: https://arxiv.org/abs/1303.3997v2

129. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009; 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

130. Quinlan AR, Hall IM. Bedtools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

131. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29: 24–26. https://doi.org/10.1038/nbt.1754 PMID: 21221095

132. Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, et al. The iPlant collaborative: Cyberinfrastructure for enabling data to discovery for the life sciences. PLoS Biol. 2016; 14: e1002342. https://doi.org/10.1371/journal.pbio.1002342 PMID: 26752627