Review Article

Centromeres: From chromosome biology to biotechnology applications and synthetic genomes in plants

Jingwei Zhou1,†, Yang Liu2,†, Xianrui Guo3, James A. Birchler4,*,†, Fangpu Han2,† and Handong Su1,5,†

1National Key Laboratory of Crop Genetic Improvement, Hu Bei Hongshan Laboratory, Shenzhen Institute of Nutrition and Health, Huazhong Agricultural University, Wuhan, China
2State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China
3Laboratory of Plant Chromosome Biology and Genomic Breeding, School of Life Sciences, Linyi University, Linyi, China
4Division of Biological Sciences, University of Missouri, Columbia, Missouri, USA
5Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

Received 9 May 2022; revised 13 June 2022; accepted 15 June 2022.
*Correspondence (Tel 573-882-4905; fax 573-882-0123; email birchlerj@missouri.edu (J.A.B.); Tel +86-27-87282130; fax +86-27-87384670; email shandong@hzau.edu.cn (H.S.); Tel +86-10-64807926; fax +86-10-74854467; email fphan@genetics.ac.cn (F.H.))
†These authors contributed equally to this work.

Keywords: centromere, CENH3, CRISPR/Cas, haploid and polyploid, chromosome engineering, synthetic genome.

Summary

Centromeres are the genomic regions that organize and regulate chromosome behaviours during cell cycle, and their variations are associated with genome instability, karyotype evolution and speciation in eukaryotes. The highly repetitive and epigenetic nature of centromeres were documented during the past half century. With the aid of rapid expansion in genomic biotechnology tools, the complete sequence and structural organization of several plant and human centromeres were revealed recently. Here, we systematically summarize the current knowledge of centromere biology with regard to the DNA compositions and the histone H3 variant (CENH3)-dependent centromere establishment and identity. We discuss the roles of centromere to ensure cell division and to maintain the three-dimensional (3D) genomic architecture in different species. We further highlight the potential applications of manipulating centromeres to generate haploids or to induce polyploids offspring in plant for breeding programs, and of targeting centromeres with CRISPR/Cas for chromosome engineering and speciation. Finally, we also assess the challenges and strategies for de novo design and synthesis of centromeres in plant artificial chromosomes. The biotechnology applications of plant centromeres will be of great potential for the genetic improvement of crops and precise synthetic breeding in the future.

Introduction

The term centromere was first applied to entities at chromosome sites connected with spindle microtubules about 90 years ago (Darlington, 1936). Since their initial characterization, centromeres have long been considered as the most mysterious parts of the genome. Over the last 50 years, many efforts have been made to investigate the highly repetitive DNA components of centromeres in different species from yeast, animals, to plants (Fukagawa and Earnshaw, 2014). The epigenetic nature of centromeres was revealed gradually after the discovery of the centromere-specific histone variant H3 (CENH3) and related regulatory factors in a variety of eukaryotic organisms (Black and Cleveland, 2011; Earnshaw, 2015; Earnshaw and Migeon, 1985; Earnshaw and Rothfield, 1985). The establishment, maintenance and propagation of centromeres were determined by CENH3 nucleosomes with a tight regulatory mechanism (McKinley and Cheeseman, 2016). The conformation of centromeric chromatin changes dynamically during the cell cycle, indicating a spatial and temporal regulation of centromere function (Blower et al., 2002; Liu et al., 2017; Muller and Almouzni, 2017; Su et al., 2017). Currently, the centromere is generally accepted as the unique chromatin structure that mediates the proteinaceous macromolecular kinetochore formation, which is essential for chromosome segregation and genomic stability (Figure 1a; Kursel and Malik, 2016; Westhorpe and Straight, 2016). Novel epigenetic markers, including threestranded R-loop structures, noncoding RNAs and N6-adenine methylation (6 mA) modifications, were observed in centromere repeat sequences (Kabeche et al., 2018; Liang et al., 2018; Liu et al., 2021b). They were involved in the loading of CENH3 nucleosomes, assembly of the kinetochore and proper chromosome segregation.

With the rapid advance of single molecule sequencing technology, an exciting time for centromere research has arrived. Detailed genomic and epigenetic maps of centromeres were completed under the ‘telomere-to-telomere’ (T2T) genome references both in plants and in animals (Altemose et al., 2022;
These results revealed rapid and ongoing variations of centromere repetitive sequence, as well as their role in the context of chromosome biology, genome stability, disease and speciation (Miga and Alexan- drov, 2021). Plentiful tools, including the CRISPR/Cas and synthetic genome systems, are now available to explore the genetic and epigenetics basis of centromere function (Luo et al., 2018; Shao et al., 2018; Yadav et al., 2020). Here, in light of this great progress, we review current advances in understanding of centromere composition, structure and evolution in plants. We highlight the regulatory mechanisms of centromere formation, and the role of centromeres in regulating chromosome behaviour during cell cycle. Additionally, the review will discuss the biotechnical application of centromere manipulations to induce ploidy changes and chromosome engineering for crop improvement and breeding. In the field of plant speciation and evolution, we prospect ways to overcome the bottleneck of centromere research and suggest some future strategies in the era of multicellular synthetic genome biology.

The DNA composition and structure of centromere in plant

Centromeres are located in the primary constriction of a chromosome. For most species, they usually contain highly repetitive DNA sequences, which are largely missing from the initial reference genome (Kursel and Malik, 2016). Tandemly-arrayed-satellite repeats (TR) and centromeric specific retroelements (CR) in most plant species. The repetitive sequences are wrapped around the centromeric specific histone H3 variant (CENH3). CENH3 nucleosomes are stringently regulated and determine the establishment, maintenance and propagation of centromeres. The outside of the centromeres consist of a proteinaceous macromolecular kinetochore complex. Kinetochore architecture includes a constitutive centromere associated (CCAN) network and the outer KMN (KNL1-MIS12-NDC80) network. Kinetochore regulators such as spindle assembly checkpoint (SAC) and chromosomal passenger complex (CPC) are involved in kinetochore assembly and its association with outermost microtubule. (b-d) DNA (b), chromatin (c) and kinetochore (d) compositions of maize centromere. Maize centromere-specific satellite repeat (CentC) are labelled in green (b). Immunostaining of CENH3 (c) and NDC80 (D) signals (red) on chromosomes. Blue indicates chromosomes counterstained with 4’, 6-diamidino-2-phenylindole (DAPI).

Figure 1 DNA, chromatin and kinetochore compositions of plant centromeres. (a) Centromere is composed of highly repetitive DNA sequences. Specific CENH3 chromatin defines centromere function, and kinetochore proteins assemble based on CENH3 chromatin. Centromeric DNA sequences includes the tandemly-arrayed-satellite repeats (TR) and centromeric specific retroelements (CR) in most plant species. The repetitive sequences are wrapped around the centromeric specific histone H3 variant (CENH3). CENH3 nucleosomes are stringently regulated and determine the establishment, maintenance and propagation of centromeres. The outside of the centromeres consist of a proteinaceous macromolecular kinetochore complex. Kinetochore architecture includes a constitutive centromere associated (CCAN) network and the outer KMN (KNL1-MIS12-NDC80) network. Kinetochore regulators such as spindle assembly checkpoint (SAC) and chromosomal passenger complex (CPC) are involved in kinetochore assembly and its association with outermost microtubule. (b-d) DNA (b), chromatin (c) and kinetochore (d) compositions of maize centromere. Maize centromere-specific satellite repeat (CentC) are labelled in green (b). Immunostaining of CENH3 (c) and NDC80 (D) signals (red) on chromosomes. Blue indicates chromosomes counterstained with 4’, 6-diamidino-2-phenylindole (DAPI).
of most centromeric satellite repeats range from 150 to 180 bp, for example, 180-bp pAL1 repeat in Arabidopsis thaliana, 156-bp CentC repeat in Zea mays and 155-CentO in Oryza sativa (Birchler and Han, 2009). However, several kilobase- to megabase-sized satellite unit were also detected in centromeres of several potato (Solanum tuberosum) and common wheat (Triticum aestivum) chromosomes (Gong et al., 2012; Su et al., 2019). Recently, a study in the maize B chromosome revealed that the B centromere-specific repeats share analogous centromere organizations with that of A centromeres (Blavet et al., 2021). Higher-order repeat (HOR) structure of centromeric satellite repeats, that has been previously described in primate, is widespread across the plant kingdom (Melters et al., 2013). Holocentromeres with the kinetochores extended along almost the entire chromosome were also detected in some plants (Steiner and Henikoff, 2014). Centromere-specific satellite arrays were found to be associated with CENH3 nucleosomes genome-wide and interspersed among euchromatin in holocentromeres of the plant Rhynchospora pubera (Marques et al., 2015). These results suggest that complex diversity in centromere composition and organization is generated during the process of speciation and evolution, in conflict with their conserved function (Henikoff et al., 2001; Kurse and Malik, 2018).

How did these highly repetitive sequences evolve, and what were the biological significance of centromere variations? A significant level of genetic variations were exploited within centromeric repeats of A. thaliana, and the core centromeres of Arabidopsis are found to be made of a specific, highly homoneous satellite type (Maheshwari et al., 2017). These results were also confirmed in the recently released complete Arabidopsis genome with ultra-long-read DNA sequencing technology (Naish et al., 2021). Similar high levels of sequence polymorphism within CentC copies and stable CENH3 binding were detected in maize and its wild Tripsacum relatives (Gent et al., 2017).

Subgencode-abundant centromeric specific satellites are located in the centromere of B- and D-subgenome of common wheat, and display increasing sequence identity during the evolution from the diploid and tetraploid to hexaploid wheat (Su et al., 2019). The homogenization and diversification of centromere repeats suggest two opposing forces for centromere evolution in plants (Naish et al., 2021). Assays of centromeres in 23 maize inbreds strongly suggest that inbreeding is a major driver of centromere DNA replacement, favoured by post-domestication selection for centromere-linked genes likely affecting key domestication or agricultural traits (Schneider et al., 2016). Hence, multilayered mechanisms may collaboratively drive the evolution of these highly repetitive sequences.

Furthermore, the variation of centromeres have been associated with loss-of-centromere function and often result in genomic alterations or instability and diseases, particularly in human (Aldrup-MacDonald et al., 2016; Barra and Fachinetti, 2018). Heterogeneous alterations in centromeric and pericentromeric sequences were detected in tumour cells (Saha et al., 2019). Reports from human cells demonstrated that the specific inactivation of centromere generates individual chromosome segregation errors, and trigger a broad spectrum of chromosome shattering that recapitulate genomic features of human cancer disease (Ly et al., 2017; Ly et al., 2019). The ‘T2T’ reference genomes will cast new insight on the genetic variation and selection of centromere sequences, and the consequences of centromere variations will be better revealed in plants. The molecular mechanisms that drive centromere variation and evolution could be potentially exploited to unlock genetic linkages within centromere regions for crop improvement in the future.

Epigenetic regulation of centromere formation and maintenance in plants

Despite the contributions from the DNA repeat sequences, centromeres identity and function are defined epigenetically by the presence of the centromere-specific histone H3 variant, CENH3 in plants or CENP-A in animals (Dawicki-McKenna and Black, 2019; Drinnenberg et al., 2016). However, how CENH3 nucleosome convey unique biophysical properties to affect centromere chromatin is poorly understood. Bioinformatic assays reveal that a single CENH3 nucleosome is wrapped by a centromere satellite unit and usually translational-phased with periodicity on these satellite repeats in different species (Figure 1a and c), suggesting that CENH3 nucleosomes are stabilized for the evolved centromeric satellite repeats (Iwata-Otsubo et al., 2017; Su et al., 2019; Zhang et al., 2013c). Furthermore, biochemical assays reveal the higher flexibility of the repeat DNA end of CENH3 nucleosomal in human cells, and the flexible end prevents histone H1 linker binding to the CENH3 nucleosome and allows kinetochore complex assembly (Hasson et al., 2013; Roulland et al., 2016). The association properties of CENH3 nucleosomes with centromeric repeats are essential to establish critical networks for the fidelity of chromosome dynamic during cell cycle.

The deposition of CENH3 nucleosomes is stringently regulated, and their positions are dynamic during evolution in plants (Feng et al., 2020; Sandmann et al., 2017). Current cognition suggests that CENH3 nucleosomes can de novo assemble at a site distinct from original centromeres and induce kinetochore formations, following inactivation or deletion of original centromeres (Figure 2a; Fu et al., 2013; Zhang et al., 2013a). Chromosome breaks induced by pollen irradiation demonstrated the regular occurrence of both centromere birth and death following chromosomal rearrangement during a narrow development time, and a sequential series of de novo formations and inactivations of centromere were displayed in maize (Figure 2b; Liu et al., 2015a; Liu et al., 2020b). Chromatin dynamics underlying these regions before and after centromere formation suggest that CENH3 seeding was affected by original chromatin state before centromere formation, and the CENH3 loading can also reshape the chromatin state after centromere formation (Su et al., 2016; Zhang et al., 2013a). Comparative genomics has revealed common occurrences in chromosomal end-to-end fusions and insertions of one chromosome into another, following regular centromere birth and death (Birchler and Han, 2018; Liu et al., 2020b; Lysak, 2022). For example, the karyotype diversifications were found to be marked with frequent centromere repositions in the Arabidopsis crucifer tribe (Figure 2c). Synteny conservation assays suggest that dynamic turnover of centromeres drives karyotype evolution and diversity through chromosome breakage and centromere inactivation in closely related Drosophila and Malassezia species, respectively (Bracewell et al., 2019; Sankaranarayanan et al., 2020). Furthermore, centromere expansion were also observed in maize chromosomes after transfer to the oat background (Figure 2d; Mandakova et al., 2020; Wang et al., 2014). Much work is needed to further explore the molecular mechanism and biological effect of CENH3 loading, spread and centromere dynamic on chromosomes during long-term evolution in plants.

Above-mentioned results show that the repetitive DNA sequence is neither necessary nor sufficient for centromere
centromeric polymerase II from late mitosis to early G1 (Chan and Henikoff, 2018). In maize, centromere-encoded RNAs are dicentric chromosome was found with only a single primary constriction, suggesting the inactivation of one centromere. (c) Centromere Reposition. De novo centromere formation in a different position on the same chromosome, and inactivation of the original centromere during genome evolution. (d) Centromere Expansion. The chromosomes of maize are much smaller than those of oat. When the maize chromosomes are transferred to the oat, the size of maize centromeres expand in adaptation to the oat genome environment.

function. However, the function of centromere DNA composition and their arrangement remain unclear. Evidence is accumulating that the transcripts and/or transcriptions from centromere repetitive sequences play important roles to maintain centromere chromatin and assemble kinetochores (Liu et al., 2021a; Talbert and Henikoff, 2018). In maize, centromere-encoded RNAs are found to be an integral component of the centromere, and the binding of CENP-C with DNA is stabilized by single-stranded RNAs (Du et al., 2010; Topp et al., 2004). Studies in human cells found that centromeric α-satellite repeats are transcribed actively by RNA polymerase II from late mitosis to early G1 (Chan et al., 2012). Depletion of these centromeric transcripts results in the mislocalization of centromere proteins, which in turn influences segregation of chromosomes during cell division (Blower, 2016; Rosic et al., 2014). A recent study revealed that point centromere activity requires an optimal level of centromeric noncoding RNAs in yeast (Yick Hin Ling and Yuena, 2019). Non-canonical secondary DNA (non-B-form) structures are found to be enriched at the centromeres of several species (Kasinathan and Henikoff, 2018; Patchigolla and Mellone, 2022), which may facilitate their transcription and contribute to centromere specification. Circular RNAs derived from centromeric CRM1 retrotransposon were found to bind with centromere through RNA:DNA hybrids (R-loop, a non-B-form DNA structure) in maize, thereby promoting centromeric chromatin loop formations to regulate proper CENH3 localization (Liu et al., 2020a). The further dissection of R-loops with ssDRIP reveals the R-loop structure is suitable for CENH3 nucleosome loading in maize (Liu et al., 2021b). R-loops are also found to be associated with maintaining centromere stability and identity in budding yeast and human cells (Kabeche et al., 2018; Mishra et al., 2021; Racca et al., 2021). These functional studies reveal a new chapter in centromere biology and genome stability. However, it is still unclear what the molecular mechanisms result in centromere transcription, especially since the transcription start sites are challenging to detect within these repetitive sequences.

Role of centromeres for dynamic chromosome behaviours in mitosis and meiosis

The centromere is the hub of the chromosome and mediates the assembly of macromolecular kinetochore complexes that mark the sites for microtubule attachment (Figure 1d). Kinetochore defects will lead to chromosomal instability (CIN), and many human diseases in tumours and syndromes. Kinetochores are highly dynamic and their assembly is regulated by the cell cycle, which ultimately ensures proper chromosome segregation during both mitosis and meiosis (Hara and Fukagawa, 2018; McKinley and Cheeseman, 2016). More than 100 kinetochore proteins have been identified from yeast to mammalian cells, ranging from the outer kinetochore Knl1-Mis12-Ndc80 (KMN) network to the inner constitutive centromere associated network (CCAN) and some regulatory proteins (Lampert and Westermann, 2011). CCAN and KMN complexes physically connect the centromeric chromatin and the spindle microtubules, respectively (Borek et al., 2020; Pesenti et al., 2016). Kinetochore proteins have a great diversity in eukaryotes, and gene duplications played a major role in shaping differing kinetochore architectures (Tromer et al., 2019; van Hoolf et al., 2017). The homologues of the KMN network have been identified in plants including Ndc80, Nufl2, Spc24, Knl1, Mis12 and Nnf1 (Allipra et al., 2021; Du and Dawe, 2007; Li et al., 2021; Li and Dawe, 2009; Shin et al., 2018; Su et al., 2021). They display conserved functions involved in microtubule organization and chromosome segregation through cell division, and the mutations cause defective developmental phenotypes in plants. However, some kinetochore proteins are missing in plant species, and their molecular architectures and co-evolutionary regulatory patterns are distinct in plants (Komaki and Schnittger, 2016; van Hooff et al., 2017). The extensive diversity suggests the mosaic origin of eukaryotes kinetochores and evolutionary flexibility of essential cellular process.

In addition to the constitutive kinetochore proteins, the functions of many centromere regulatory factors are revealed recently in plants. The spindle assembly checkpoint (SAC) signal is the conserved regulatory mechanism in centromeres that control the cell cycle and genome stability (Lara-Gonzalez et al., 2012). BFM1, one of the SAC components, phosphorylates histone H2AThr133 to regulate chromosome alignment and segregation in maize (Su et al., 2017). Temporal and spatial dynamics with the positions of centromere nucleosomes were necessary for proper chromosome orientation and segregation through the cell cycle (Figure 3a; Liu et al., 2017; Su et al., 2017). The plant spindle assembly checkpoint system has a less efficient
checkpoint function and evolved various other cellular activities (Komaki and Schnittger, 2017). Neo-functions of these regulatory factors were associated with the kinetochore diversity generated by duplication of kinetochore genes in plants. A recent study in Arabidopsis revealed that Bub3 functions in microtubule reorganization signalling and development of phragmoplasts during cytokinesis (Zhang et al., 2018; Zhang et al., 2021). The understanding of kinetochore function in plants is relevant for future breeding, genomics and evolutionary studies.

On the contrary, growing evidence confirms that centromeres are central to chromosome behaviour during meiotic cell division (Figure 3b). Detailed cytogenetics revealed that centromeres form pairwise associations at the leptotene stage, preceding the formation of the telomere bouquet and in early prophase I of meiosis in maize (Zhang et al., 2013b). Non-homologous centromere clustering and/or coupling seems to occur as an early step, and precede centromere pairing of homologous chromosomes (Obeso et al., 2014). This process appears to be a common feature in a number of species and may play a fundamental role to facilitate chromosome pairing in meiosis (Da Ines and White, 2015). Functional studies suggest that centromere activity and subunits of the cohesin complex were to participate in centromere pairing and to stabilize meiotic homologue chromosome pairing in different species (Hatkevich et al., 2019; Zhang et al., 2020). A structural reorganization of the centromeric chromatin and kinetochores coincides with key events during early meiosis of synopsis in different species (Borek et al., 2020; Sepsi et al., 2017).

In addition to the role in chromosome pairing, centromeres serve as the basis to adopt special geometry and force for meiosis I sister kinetochore co-orientation, of which the sister kinetochores of a chromosome are captured by the microtubules from the same spindle pole (Prosee et al., 2020; Watanabe, 2012). Mechanically fused sister kinetochores were reported with the aid of monopolin complex during meiosis I in yeast (Sarangapani et al., 2014). Furthermore, a meiosis-specific kinetochore protein (Meikin), the functional analogous of monopolin complex, was identified to be a conserved regulator to protect centromeric

Figure 3  Centromere behaviour during mitosis, meiosis cell cycle and interphase. (a) Centromere roles underlying the dynamic chromosome behaviour during mitosis. During prophase, the phosphorylation of histone H2AT133 (H2AT133ph) occurs in the CENH3 nucleosome. Histone H3T3 phosphorylates (H3T3ph) throughout the entire chromatin, and these two marks mix with CENH3 nucleosomes within centromeres. In anaphase, CENH3 and H2AT133ph nucleosomes occupy the outer centromere, while H3T3ph nucleosomes locate in the inner centromere. The kinetochore protein complex formed by the CCAN and KMN network connects centromeres and microtubules. The kinetochore proteins identified in plants are shown. (b) Centromere roles for dynamic chromosome behaviours during meiosis. At the end of premeiotic S phase, the cohesin complex protects sister chromatids along the entire chromosome, and homologous centromere clustering and pairing occurs in leptotene stage. The MIS12–NDC80 bridge fuses the sister kinetochores to initiate mono-orientation during meiotic metaphase I in plant. (c) Centromere architecture in the three-dimensional genome architecture in plants. The Rabl configuration: Centromeres and telomeres occupy opposite sides of the nucleus, represented in wheat, rye, barley and oat. Non-Rabl configuration, where centromeres and telomeres are scattered around the periphery of the nucleus, represented in sorghum and maize. The Rosette-like configuration occurs when the chromosomes of centromeric and pericentromeric regions are tightly packed into chromocenters and anchor proximal euchromatin loops, represented in Arabidopsis.
cohesion and facilitates mono-orientation in mammals (Kim et al., 2015). A visible MI12–NDC80 bridge fuses sister kinetochores to regulate sister kinetochore co-orientation and initiate the reductive division during meiosis I in maize (Li and Dawe, 2009). Misorientation of kinetochores during meiosis I was observed in cdk20.1 (Cell Division Cycle 20) mutant of Arabidopsis thaliana, suggest its role for kinetochore orientation during meiosis (Niu et al., 2015). Further understanding of conformation changes at centromeric regions during the early phase of meiosis I will be of great significance for the field of plant reproduction and development (Bar and Amon, 2008), and thus can promote the application in agricultural breeding.

Centromere clustering in three-dimensional genome architecture in plants

In addition to these well-established roles, centromeres were reported recently to have a strong impact on the three-dimensional genome architecture (3D) and chromatin regulation (Figure 3c). Previous cytological observations in root-tip cells of wheat, rye, barley and oat with larger genome sizes suggest Arabidopsis thaliana, suggest its role for kinetochore orientation (Nishimura et al., 2018). These results suggest that centromere clustering is based on the function of centromere and is independent of the chromosomal DNA context. Another work in maize shows that circular RNAs derived from centromeric retrotransponson (CRM1) form R-loops, and prompt centromeric chromatin remodeling, function and maintenance (Bobkov et al., 2018). Overall, because the centromere influences the genome-wide chromosomal interactions, it may prevent certain contacts between particular chromosomal or intra-chromosomal regions (Muller et al., 2019). Failure in centromere clustering results in partial defects in the silencing of repetitive elements, indicating that a proper centromere distribution is required to maintain the constitutive heterochromatic state around centromeres (Fadeken et al., 2013).

Manipulation of centromeres induces ploidy changes in plants

Centromeres are associated with faithful chromosome segregation during the cell division process. The dysfunction of centromeres induces chromosome instability, aneuploidy or cytokinesis failure, and can lead to developmental delays or cancer in animals. However, the consequence often leads to frequent ploidy changes in plants, including haploid and polyploid formation (Keceli et al., 2020; Kozgunova et al., 2019; Miga and Alexandrov, 2021).

Uniparental genome elimination occurs frequently in the interspecific crosses of plants and results in uniparental haploid progeny naturally, for example, in wide crosses such as wheat × maize, H. vulgaris × H. bulbosum and wheat × Pennisetum glaucum (Ishii et al., 2016). These crosses are used to produce doubled haploid plants to accelerate the crop breeding process. Centromere deficiency is thought to play a centre role in the uniparental genome elimination (Figure 4a). Failure of CENH3 incorporation into centromeres was shown to precede uniparental genome elimination (Sanei et al., 2011) and to eventually lead to complex chromosome rearrangement during the early embryo development in Arabidopsis and in barley (Comai and Tan, 2019). Initial pioneering proof in Arabidopsis indicated that crossing plants expressing different forms of CENH3, results in haploids induction frequencies as high as 45% (Ravi and Chan, 2010). Point mutations and small deletions of CENH3 impaired its loading and induced haploid plants at similar frequencies in Arabidopsis (Karimi-Ashiyani et al., 2015; Kuppu et al., 2015; Kuppu et al., 2020). The manipulation of CENH3 in wheat and maize have been used to generate haploids (Lv et al., 2020; Wang et al., 2021). Recently, parental biased removal and loading of CENH3 and other kinetochore proteins was revealed to result in epigenetically distinct centromeres that initiate a strong mating barrier and produce haploids in plant (Marimuthu et al., 2021), concerning that the CENH3 protein is a player in planta haploid induction. The conserved CENH3 and other kinetochore proteins provide potential targets for diverse crop application of centromere-induced haploid induction.

In addition to haploid generation, the disruption of centromeric proteins may also lead to the generation of polyploids in plants (Figure 4b). The knock-down of several kinetochore components results in chromosome mis-segregation and cytokinesis failure, and developed into somatic polyploidy cells in moss Physcomitrella patens (Kozgunova et al., 2019). Prolonged SAC activation in Arabidopsis caused a reset of the cell cycle producing duplicated chromosomes, but no nuclear division (Komaki and Schnittger, 2017). It has become clear that whole-genome duplications regularly occurred during the evolution of the plant kingdom (Wendel et al., 2016). As the kinetochores have important implications for genome stability and plant evolution, the ploidy levels can be readily made in plants. However, the mechanisms that are associated with ploidy changes when centromere proteins are altered are still mysterious. Their
Elucidation is likely to generate multiple applications in agricultural breeding for the future. Centromere-targeted chromosome engineering for novel chromosomes and even new species formation

As fragile sites of chromosomes, centromeres are susceptible to breaks caused by chromosome mis-segregation, incorrect DNA replication and recombination, and improper centromere topology (Barra and Fachinetti, 2018). The loss, gain or repositioning of centromere occurred frequently, and influence chromosome number and rapid karyotype turnover during plant genome evolution, which is thought to be important in speciation (Wendel et al., 2016). Frequent centromere variation was observed in wheat aneuploids and its wild hybrids, resulting in complex chromosomal reorganizations and novel chromosome formations, which in turn act as a reproductive barrier and facilitate speciation (Guo et al., 2016). In fact, given that the centromere is largely repetitive sequences up to megabases in size, it provides many potential targets for manipulation with genome editing tools for chromosome engineering. Recently, the editing of centromere tandem repeats proved to be a valuable tool for organ-specific cell elimination in Arabidopsis (Schindele et al., 2022). Hence, centromere editing made possible by the advent of genome editing tools, will be an important strategy in the generation of novel chromosomes, which will in turn accelerate karyotype change for crop improvements.

Combined with currently available tools, it should be possible to mimic natural chromosomal rearrangements and to create new chromosomes or plant species by changing centromere positions and chromosome numbers (Figure 5). Simultaneous double-strand breaks (DSBs) at multiple centromeric repeats were targeted with CRISPR and led to chromosome shuffling in Cryptococcus neoformans. Karyotype shuffled strains exhibited severe defects in sexual reproduction with the parental genotype.

Figure 4  Centromeres manipulations induce ploidy changes in plants. (a) Centromere-mediated haploid induction. cenh3 mutants can induce haploids when crossing with CENH3 wild-type line. During mitosis of the hybrid zygote, the CENH3 nucleosome can incorporate in the chromosome of WT normally, while CENH3 nucleosomes cannot load on the chromosome from the cenh3 mutant lines, whose centromeres lose the ability to bind with microtubules. Following anaphase of hybrid zygotes, these laggard chromosomes and form micronuclei are torn into fragments, eventually leading to genome elimination and haploids. (b) Centromere-mediated somatic polyploidization in plants. Knockout of the kinetochore proteins (cenp-c, cenp-x, ndc80 and knl1) in Physcomitrella patens and SAC (bmf2, bmf3, cdc20 and bub3;3)-related proteins in Arabidopsis leads to weaker binding of the kinetochore complex to microtubules during mitosis, resulting in cytoplasmic failure and the generation of somatic polyploidy plants.
providing experimental evidence supporting models in which centromere-mediated chromosomal rearrangements reshape eukaryotic genomes and may cause reproductive isolation and subsequently speciation (Yadav et al., 2020). Mutations in inner kinetochore components induce centromere repositioning in fission yeast, generates a reproductive barrier, suggesting a functional role of evolutionary new centromeres in speciation (Lu and He, 2019). Selected centromere inactivation causes chromosome-specific segregation errors, and drives extensive structural rearrangement or abnormalities that are associated with genomic instability in common human diseases (Ly et al., 2017; Ly et al., 2019). S. cerevisiae cells with only a single- or two-chromosomes were produced via CRISPR/Cas-mediated chromosome engineering, and it remodelled the genome architecture to produce reproductive isolation (Luo et al., 2018; Shao et al., 2018). CRISPR/Cas-mediated chromosome inversions and translocations have been demonstrated in plants, and proposed for accelerating crop breeding (Beying et al., 2020; Ronspies et al., 2021; Schmidt et al., 2020; Schwartz et al., 2020). Restructuring and reshuffling chromosomes would answer the basic questions on how different centromere positions on a chromosome might influence chromatin state and gene expression, which may lead to develop new crop varieties with improved phenotypes.

**Overcoming the centromere obstacle in synthetic plant genomes**

Synthetic genomics is a new paradigm to dissect the nature of chromosomes and to engineer novel biological functions, providing a promising system for applied science (Schindler et al., 2018). Entire genomes of several viruses and bacteria have been chemically synthesized and applied to drive normal cell processes during the last decades (Cello et al., 2002; Gibson et al., 2008; Gibson et al., 2010; Hutchison 3rd et al., 2016). The first eukaryotic genome synthesis project on yeast (Sc2.0) is ongoing, and about half of the genome has been synthesized and functionally tested (Annaluru et al., 2014; Shen et al., 2017; Wu et al., 2017; Xie et al., 2017; Zhang et al., 2017). The full design of a synthetic yeast genome at the genome level provides a ‘grammar’ of eukaryotic life (Richardson et al., 2017). Due to its complexity, this bottom-up approach has not yet been planned to synthesize a chromosome in plants. With the development of synthesis technology and gradual reduction of costs, there has been a recent surge of interest in applying these principles to plant genome construction. However, the detailed sequence organization and structure of centromeres are still largely unknown in most plant systems. Furthermore, kinetochore proteins do not associate with natural centromere sequences introduced in plants by transformation (Phan et al., 2007). Chemical synthesis of these highly repetitive centromere sequences also presents a major challenge. The centromere obstacle therefore hampers the process of synthetic artificial chromosome in plants (Birchler et al., 2016; Dawe, 2020).

**De novo centromeres were generated with the removal of native centromeres in plants** (Liu et al., 2015b; Su et al., 2016). This suggests that the deposition of CENH3 nucleosomes is potentially determined by its pre-existing presence at a site and not by DNA sequence in plants, which is quite different from the centromeres in yeast that were determined by DNA sequence (Morris and Moazed, 2007). The ‘de novo strategy’ can be applied towards the creation of synthetic chromosomes in higher organisms (Figure 6a). For example, human artificial chromosomes were recently developed where initial CENP-A nucleosomes are seeded on a non-repetitive sequence based on the LacO-LacI-HJURP tethering system, and this native-centromere-bypass approach facilitates mammalian synthetic genome efforts (Logsdon et al., 2019). Top-down approaches were also utilized for the construction of plant artificial chromosomes based on telomere-mediated chromosome truncation with endogenous centromeres that can be used as engineered chromosomes.

**Figure 5** Centromere-targeted chromosome engineering for new chromosomes, and new species formation in plants. (a) Simultaneous editing of the repeat sequences flanking the centromere in one chromosome (from top to bottom, orange-green) by CRISPR/Cas system induces two double-strand breaks (DSBs). During the DNA repair process, intra-chromosomal translocations occurred with a pericentric inversion on the chromosome (top to bottom, green-orange). (b) Multiple DSBs can be generated via the action of the CRISPR/Cas system at centromere-specific repeats of different chromosomes, resulting in the formation of inter-chromosomal translocations, and reshuffling of the genome karyotype. (c) Two DSBs can be generated simultaneously using the CRISPR/Cas system on the pericentromere of different chromosomes. The cut ends fuse to form a dicentric chromosome and a chromosome fragment without an original centromere. Following this, one centromere on the dicentric chromosome can be inactivated, and of a de novo centromere can be formed on the acentric chromosome fragment.
(Figure 6b). This method has been achieved with some mini-chromosomes in various plant species (Birchler et al., 2016). With the advent of the ‘telomere to telomere’ genome era, the comprehensive understanding of the (epi)genetic nature of centromeres at the genomic level will accelerate the development of synthetic genome biology in plants. The final destination towards fully synthetic plant chromosomes may involve a mixture of these two strategies.

Conclusions and prospects

Centromeres are the specialized chromosomal regions that are required for the maintenance of genomic stability during cell divisions, and their inheritance and function are strictly dependent on genetic and epigenetic components in most species. In the present review, we summarize the recent progress involved in the molecular mechanisms that are associated with centromere formation and function in plants.

As presented, centromeres have highly diverse DNA compositions and structures of the outer kinetochore complex. This in turn influences their potential biological functions, which are associated with important processes such as chromosome biology, genome instability and human disease. In our previous studies, the regular occurrence of both centromere birth and death following chromosomal rearrangement illustrated that epigenetic factors play important roles in centromere maintenance and high-order chromatin structure. Centromere transcription and Non-B form DNA (R-loops) have been suggested to participate in the modulation of centromere establishment, maintenance and propagation. The phosphorylation of histone H2AThr133, H3Thr3 and CENH3 nucleosomes in centromere regions and their spatial–temporal dynamics suggest vital roles of centromeres in the proper chromosome orientation and separation during the cell cycle (Figure 3a). These results not only provide new knowledge for understanding the function, evolution and speciation, aspects of centromeres in eukaryotes, but also potential applications towards artificial genetic engineering.

Recently, centromere-mediated ploidy changes have been applied for accelerating breeding and crop improvement. In the near future, more research should be conducted to evaluate the molecular mechanisms that drive haploid and polyploid induction when centromere proteins are disrupted. Precise genome editing of CENH3, and other identified kinetochore proteins in plants will improve the efficiency of these ploidy changes. CRISPR genome editing technology can now effectively induce heritable chromosomal rearrangements, such as inversions and translocations in...
plants. Cumulative evidence has demonstrated that centromere editing might unlock genetic linkages within centromere regions and thus facilitate precise chromosome engineering in crop improvement. Centromere-mediated chromosome reshufflings and centromere repositionings can generate novel chromosomes, which may induce reproductive barriers and facilitate speciation. The study of chromosome reshuffling on gene expression regulation and three-dimensional chromatin architecture will help elucidate the rapid establishment, stability and evolution of novel plant genomes.

Furthermore, synthetic genomes are a promising tool to direct future crops with new properties. The construction of synthetic plant chromosome, in particular, likely relies on a basic mechanistic understanding of centromere structure and function. The complexity and highly repetitive nature of the centromere poses a significant hurdle for its functional synthesis. It is still a great challenge to chemically synthesize several kilobase-megabases of centromere repetitive sequences, and current transformation methods are not sufficient to deliver in-vitro-synthetic chromosomes or chromosome fragments into plant cells. With the rapid development of genome sequencing, synthesis technologies, genome editing and nanoparticle transformation tools from cross-disciplinary collaborations will offer important potential approaches to facilitate fully synthetic plant genomes.

Acknowledgements
This project was financially supported by the National Natural Science Foundation of China (32170571 and 31991212). Funding was also provided by the National Key Research and Development Program of China (2021YFF1000800), the Fundamental Research Funds for the Central Universities (2021KPY008) and Natural Science Foundation of Hubei Province of China (2021CB097). We thank the high-performance computing platform at National Key Laboratory of Crop Genetic Improvement in Huazhong Agricultural University.

Conflict of interest
The authors declare no conflict of interest.

Author's contributions
All authors contributed to the writing of this review article. H.D.S. F.P.H. and J.B. discuss the manuscript. H.D.S., J.W.Z., Y.L. and X.R.G. wrote the first draft of the manuscript. F.P.H., J.B. and H.D.S. revised the manuscript. J.W.Z. prepared the figure. All authors approved the manuscript.

References
Aldrup-MacDonald, M.E., Kuo, M.E., Sullivan, L.L., Chew, K. and Sullivan, B.A. (2016) Genomic variation within alpha satellite DNA influences centromere location on human chromosomes with metastable epialleles. Genome Res. 26, 1301–1311.
Allipr, S., Anirudhan, K., Shivanandan, S., Ragunathan, A. and Maruthachalam, R. (2021) The kinetochore protein NNF1 has a moonlighting role in the vegetative development of Arabidopsis thaliana. Plant J. 109, 1064–1085.
Altemose, N., Logidson, G.A., Izbakde, A.V., Sidhwan, P., Langley, S.A., Caldas, G.V., Hoyt, S.I. et al. (2022) Complete genomic and epigenetic maps of human centromeres. Science, 376, eabi4178.
Logdon, G.A., Gambogi, C.W., Liskovkh, M.A., Barrey, E.J., Larionov, V., Miga, K.H., Heun, P. et al. (2019) Human artificial chromosomes that bypass centromeric DNA. Cell, 178, 624-639.e19
Lu, M. and He, X. (2019) Centromere repositioning causes inversion of meiosis and generates a reproductive barrier. Proc. Natl. Acad. Sci. USA, 116, 21580-21591.
Luo, J., Sun, X., Cormack, B.P. and Boeke, J.D. (2018) Kanotype engineering by chromosome fusion leads to reproductive isolation in yeast. Nature, 560, 392-396.
Lv, Y., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., Wang, Y. et al. (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. Nat. Biotechnol., 38, 1397-1401.
Ly, P., Brunner, S.F., Shoshani, O., Kim, D.H., Lan, W., Pyntikova, T., Flanagan, A.M. et al. (2019) Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. Nat. Genet., 51, 705-715.
Ly, P., Teitz, L.S., Kim, D.H., Shoshani, O., Skalitsky, H., FachiNetti, D., Page, D.C.O. et al. (2017) Selective Y centromere inactivation triggers chromosome shattering in murine and repair by non-homologous end joining. Nat. Cell Biol., 19, 68-75.
Lysak, M.A. (2022) Celebrating Mendel, McClintock, and Darlington: on end-to-end chromosome fusions and nested chromosome fusions. Plant Cell, 34, 2475-2491.
Maheshwari, S., Ishii, T., Brown, C.T., Houben, A. and Comai, L. (2017) Lysak, M.A. (2022) Variation and evolution of centromeres. Plant Cell, 34, 602-619.
Maklein, E. et al. (2015) Comparative analysis of tandem repeats from Arabidopsis. J. Mol. Biol., 433, 278-291.
Miga, K.H. and Alexandrov, I.A. (2021) Variation and evolution of centromeres. Plant Biotechnology Journal, 19, 134-149.
Niu, B., Wang, L., Zhang, L., Ren, D., Ren, R., Copenhaver, G.P., Ma, H. et al. (2015) Arabidopsis cell division cycle 20 (Cdc20) is required for Normal meiotic spindle assembly and chromosome segregation. Plant Cell, 27, 3367-3382.
Obeso, D., Pezza, R.J. and Dawson, D. (2014) Couples, pairs, and clusters: mechanisms and implications of centromere associations in meiosis. Chromosoma, 123, 43-55.
Oko, Y., Ito, N. and Sakamoto, T. (2020) The mechanisms and significance of the positional control of centromeres and telomeres in plants. J. Plant Res., 133, 471-478.
Paden, J., Mendiburo, M.J., Chlamydas, S., Schwarz, H.-J., Kremmer, E. and Heun, P. (2013) The Nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. Mol. Cell, 50, 236-249.
Pathigolla, V.S.P. and Mellone, B.G. (2022) Enrichment of non-B-form DNA at D. melangaster centromeres. Genome Biol. Evol., 14, evax054.
Pesenti, M.E., Weir, J.R. and Musacchio, A. (2016) Progress in the structural and functional characterization of kinetochores. Curr. Opin. Struct. Biol., 37, 152-163.
Phan, B.H., Jin, W., Topp, C.N., Zhong, C.X., Jiang, J., Dawe, R.K. and Parrott, W.A. (2007) Transformation of rice with long DNA-segments consisting of random genomic DNA or centromere-specific DNA. Transgenic Res., 16, 341-351.
Proese, R.F., Wenda, J.M. and Steiner, F.A. (2020) Adaptations for centromere function in meiosis. Essays Biochem., 64, 193-203.
Rabl, C. (1885) Uber zelltheilung. Gegenbaur Morphol Jahrb., 10, 214-330.
Racca, C., Britton, S., Hedouin, S., Francault, C., Calou, P. and Larminat, F. (2021) BRCA1 prevents R-loop-associated centromeric instability. Cell Death Dis., 12, 896.
Raviv, M. and Chan, S.W. (2010) Haploid plants produced by centromere-mediated genome elimination. Nature, 464, 615-618.
Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D. et al. (2017) Design of a synthetic yeast genome. Science, 355, 1040-1044.
Ronspies, M., Dorn, A., Schindele, P. and Puchta, H. (2021) CRISPR-Cas-mediated chromosome engineering for crop improvement and synthetic biology. Nat. Plants, 7, 566-573.
Rosic, S., Kohler, F. and Erhardt, S. (2014) Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J. Cell Biol., 207, 335-349.
Roulland, Y., Ouarafhni, K., Nadenov, M., Ramos, L., Shuab, M., Syd, S.H., Lone, I.N. et al. (2016) The flexible ends of CENP-A nucleosomes are required for mitotic Fidelity. Mol. Cell, 63, 674-685.
Saha, A.K., Mourad, M., Kaplan, M.H., Chefetz, I., Malek, S.N., Buckanovich, R., Markovitz, D.M. et al. (2019) The Genomic Landscape of Centromeres in Canaries. Sci. Rep., 9, 11259.
Sandmann, M., Taltbert, P., Demidov, D., Kuhlmann, M., Rutten, T., Conrad, U. and Lermontova, I. (2017) Targeting of Arabidopsis KNL2 to centromeres depends on the conserved CENPC-k motif in its C terminus. Plant Cell, 29, 144-155.
Sanei, M., Pickering, R., Kunke, N., Nasuda, S. and Houben, A. (2011) Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proc. Natl. Acad. Sci. USA, 108, E498-E505.
Sankaranarayan, S.R., Janiri, G., Coelho, M.A., Reza, M.H., Thimmappa, B.C., Sarangapani, K.K., Duro, E., Deng, Y., Alves Fde, L., Ye, Q., Opoku, K.N., Ceto, S. et al. (2014) Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J. Cell Biol., 207, 335-349.
