Seed Biology

ARTICLE

Open Access

https://doi.org/10.48130/SeedBio-2023-0024

Seed Biology **2023**, 2:24

Loss of function of *CENH3* causes genome instability in soybean

Jing Wang[1,2], Xiao-Fei Wang[1], Wei-Cai Yang[1,2*] and Hong-Ju Li[2,3*]

[1] *Center for Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*
[3] *Center for Molecular Agrobiology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*
* Corresponding authors, E-mail: wcyang@gentica.ac.cn; hjli@genetics.ac.cn

Abstract

Centromeres are the sites where kinetochores assemble and spindle microtubules anchor to the chromosomes during cell division. Centromeres are epigenetically specified by the centromeric histone H3 (CENH3). The imbalance of CENH3 loading rate or dosage on parental centromeres often leads to uniparental chromosome elimination in the offspring. A body of studies of CENH3 in genome stability have been reported in *Arabidopsis*, cotton, and many other monocots, but not in soybean (*Glycine max*), an important dicot crop. In our study, we identified a single-copy functional *CENH3* in soybean and found its role in genome stability and parent-of-origin effect caused by the mutation of a conserved glycine site and parental genetic background. This study provides evidence that knockout of *CENH3* in soybean has the potential to induce chromosome elimination and would shed light on the future development of CENH3-based haploid induction (HI) system and centromere biology in soybean.

Citation: Wang J, Wang XF, Yang WC, Li HJ. 2023. Loss of function of *CENH3* causes genome instability in soybean. *Seed Biology* 2:24 https://doi.org/10.48130/SeedBio-2023-0024

Introduction

CENH3, known as CENP-A in mammals, is a conserved centromeric histone H3 throughout the eukaryotes, which replaces the canonical H3 in nucleosomes and marks the position of centromere epigenetically[1]. CENH3 contains a variable N-terminal tail and a highly conserved C-terminal histone fold domain (HFD), with the latter folding into the nucleosomes and the former extending out of the chromosome[2]. The HFD contains four helixes and two loops, and the CENP-A targeting domain (CATD) composed of loop1 and α2 helix is sufficient for CENH3 loading onto centromeres[3]. Kinetochore proteins, such as CENP-C, are recruited to centromeres to assemble functional kinetochores by interacting with CENH3 directly or indirectly[4]. Since kinetochores are the linker between chromosomes and spindle fibers, they play an important role in elaborate chromosome mobilization during cell division. Disruption of CENH3 function interferes with kinetochore assembly and thus usually causes chromosome elimination during mitosis[5,6].

It has been found that crossing N-terminal altered CENH3 as maternal to wild-type produced 25%–45% haploids in *Arabidopsis*, and this strategy also worked in wheat and maize with haploid induction rates (HIR) of 8% and 0.86%, respectively [7−10]. Point mutations in the conserved C-terminal HFD of AtCENH3 led to 0.61%–44% haploids[11−13]. It was evidenced that impaired AtCENH3 with altered N-terminal or C-terminal point-mutations fails to load onto the centromeres from the HI parent in zygotes, leading to weak centromeres and uniparental chromosome elimination[14]. Strikingly, heterozygous +/*cenh3* in maize resulted in 5% haploids when outcrossed as maternal, but the ratio was only 0.4% in *Arabidopsis*[14,15], and specific degradation of maternal EYFP-tagged CENH3 by nanobody targeting also led to haploid production[16]. These studies indicate that the amount of CENH3 on maternal centromeres accounts for the maternal genome integrity in the offspring, and could be applied in haploid breeding[17]. CENH3 also participates in chromosome elimination during wide crosses or after replacement of AtCENH3 by its distantly homologous protein[9,18], implying its role in reproductive isolation.

Although a maternal haploid soybean has been reported in twins progenies from outcross of male sterile (*ms1 ms1*) plants and a cultivar[19], haploid induction technology in soybean has not been established. As *CENH3* in soybean has not been functionally studied yet, its identification would promote the understanding of soybean centromere assembly and development of CENH3-based haploid induction technology. In this study, using CRISPR/Cas9 gene-editing, we obtained several *Gmcenh3* mutants and confirmed the presence of a single-copy *GmCENH3* and its role in genome stability.

Materials and methods

Plant growth conditions

Glycine max varieties Williams82 (Wm82), Zhonghuang13 (ZH13) and Jack used for crosses were grown outdoors from May to October in Beijng, China. In winters, the plants were grown only for genotyping and seed propagation in Sanya, Hainan and in a greenhouse. The greenhouse was set to 28 °C, 16 h light and 8h dark.

DNA extraction and PCR

For gene cloning, genomic DNA was prepared using a simplified CTAB (cetyl trimethylammonium bromide) protocol[20]. Leaf tissues (1 cm^2) were ground in 400 μL CTAB buffer under room temperature using 0.5-mm-diameter small steel ball and tissue grinder. After the liquid was incubated at 65 °C for 1 h,

© The Author(s)

400 μL chloroform was added and the liquid was mixed gently. The mixture was centrifuged for 10 min at 12,000 rpm and 300 μL supernatant was transferred to mix with 300 μL isopropanol in a clean centrifuge tube. After incubation at −20 °C for 2 h, the liquid was centrifuged for 10 min at 12,000 rpm and the pellets were washed once with 800 μL 70% ethanol. The dried pellets were dissolved with 100 μL sterilized water.

For library preparation of genome-resequencing, genomic DNAs were prepared using a modified CTAB protocol. The leaf tissues (1 cm^2) were stored at −80 °C before use. The tissues were ground in a centrifuge tube cooled by liquid nitrogen using 0.5-mm-diameter small steel ball and tissue grinder at 30 Hz for 15 s, twice. The CTAB buffer was preheated to 65 °C in a water bath and 600 μL hot CTAB buffer was added to dissolve the tissue powder. After incubating at 65 °C for 30 min, the liquid was shortly centrifuged and 500 μL supernatant was transferred to a clean centrifuge tube. Five hundred μL chloroform: isoamyl alcohol (24:1) was added and mixed gently for 5 min. The mixture was centrifuged for 20 min at 12,000 rpm and 400 μL supernatant was transferred and mixed with 270 μL isopropanol and 30 μL 3M sodium acetate (pH 5.6) in a clean centrifuge tube. After incubation at −20 °C overnight, the liquid was centrifuged for 50 min at 12,000 rpm and the pellets were washed once with 800 μL 70% ethanol. The dried pellets were dissolved with 100 μL sterilized water and 1μL RNase was added to degrade RNA at 37 °C for 1 h. The solution was stored at −80 °C for the subsequent library preparation.

2^xEs Taq MasterMix (Dye) (Cat. CW0690H) was used in PCR-based genotyping. KOD Plus high-fidelity DNA polymerase (Lot No.7352054) was used in PCR for plasmid construction.

RNA extraction and RT-PCR

Small inflorescences with unopen flowers and young roots were cooled and stored immediately in liquid nitrogen. Total RNA isolation Kit (Vazyme Cat.RC411-01) was used for RNA extraction. FastQuant RT Kit (with gDNase) (Cat#KR106-02) was used for synthesis of the first cDNA strand.

Protein alignment

CENH3 protein sequences were aligned by ClustalW in MEGA7 and the alignment file was displayed using the online tool The Sequence Manipulation Suite (www.bio-soft.net/sms/index.html).

Plasmid construction and transformation

For subcellular localization of GmCENH3.1, GmCENH3.4 and Glyma.16G026300.2 proteins, the respective full-length coding sequences (CDSs) were cloned into the pH7WGF2 vector with enhanced green fluorescent protein (EGFP) at the N-terminus. For CRISPR/Cas9 gene-editing, two sgRNAs were designed using the online web tool CRISPOR[21]. U6 and U6-b promoters were used to drive the guide RNA oligonucleotides, respectively, and these two cassettes were cloned into the PMDC123 vector.

For transient transformation, these constructs were introduced into *Agrobacterium rhizobiaceae* strain K599 and then transformed into Wm82 seedlings as previously reported[22]. The *Gmcenh3* mutant was created by transforming CRISPR/Cas9 constructs into the Jack variety by the Beijing Dabeinong Technology Group.

Microscopy

Root tip was sliced transversely in a drop of sterilized water on a microscope slide to obtain thin sections and 0.01 mg/mL DAPI solution was added to stain the DNA for 5 min in the dark. The sections were covered by a coverslip for imaging. Imaging was performed using a Carl Zeiss LSM980 laser confocal microscope. EYFP was excited with 488 nm laser and the specific fluorescence was recorded with a 500–550 nm band-pass filter. DAPI was excited with a 405 nm laser line and the specific fluorescence recorded with a 409–490 nm band-pass filter.

Alexander staining of pollen grains

Stamens were collected right before anthesis and immersed in 100 μL Alexander's dye. After incubating at 55 °C overnight, the anthers were dissected in a drop of Alexander's dye and observed under a microscope.

Library preparation and sequencing

For PCR-free library preparation, 2 μg of DNA were used according to a PCR-free protocol BGI-NGS-JK-DNA-003. The DNA was sequenced on BGI DNBseq with paired-end 150 bp reads.

Statistical analysis

Chromosome dosage analysis

For chromosome dosage plot analyses, 150 bp single reads were mapped to the *Glycine max* reference genome sequence (GCF_000004515.6_Glycine_max_v4.0) using BWA. The software Control-FREEC (Version:11.5) was used to detect copy number variations (CNV). The window size set in this project was 50,000 bp, the sliding step size was 10,000 bp, the species chromosome ploidy was 2, and CNV detection and quantitative statistics were carried out for each sample. Python was used for data processing and python's matplotlib library was used for picture drawing. First, the data in the Ratio column is processed so that its maximum value was 3; secondly, the species analyzed in this project are diploid species, so the value of the Ratio column was then multiplied by 2 to make its maximum value 6; lastly, plot a scatter plot using the data: the X-axis was the starting position of the CNV quantity statistical area, and the Y-axis was the Ratio value. The raw data and processed data files of genome resequencing are available in the NCBI Gene Expression Omnibus under accession number GSE244481.

Molecular markers analysis

For Indels between Jack and ZH13, the raw paired-end (PE) reads from Jack were preprocessed using Cutadapt v1.18[23] to remove adaptors and trim low-quality sequences. Trimming involved in removing bases from 3' with quality values below 30, and discarding reads shorter than 80 bp. Trimmed reads were then aligned to ZH13 reference genome v2.0 (https://ngdc.cncb.ac.cn/gwh/Assembly/652/show) using BWA-MEM[24] with default parameters. Alignments were filtered for those mapped in proper pairs and above a MAPping Quality (MAPQ) of 30 using Samtools v1.14[25]. Next, Picard 2.26.10 (https://broadinstitute.github.io/picard/) was used to filter optical duplicates from experimental processes. Variants calling was performed with the HaplotypeCaller in GATK v4.1.6.0[26]. Then, only INDELs of size greater than 40 were selected and filtered (QD < 2.0, QUAL < 30.0, FS > 200.0, SOR > 10.0, ReadPosRankSum < −20.0, and ReadPosRankSum-20) using SelectVariants and VariantFiltration in GATK, respectively. To identify which INDELs in these final sets might be used to differentiate parental ancestry, snpEff v 4.3.1t-1[27] was run through for annotation and effect prediction purpose.

For Indel between Jack and Wm82, reads were aligned to Wm82 reference genome v2.0 (www.ncbi.nlm.nih.gov/datasets/genome/GCF_000004515.6/) and the HaplotypeCaller in GATK4 (Version：4.1.2.0) was used and the parameters was set as below: QD < 2.0, QUAL < 30.0, FS > 200.0, SOR > 10.0, MQRankSum < −12.5, and ReadPosRankSum-8.0.

Chi-square test

Chi-square test of segregation ratio was performed on GraphPad Prism 8.0.1. In the setting of parameters, relative risk and chi-square test were chosen.

Results

A single-copy of functional *CENH3* is encoded in the soybean genome

Two *CENH3* homologs, *Glyma.07G057300* and *Glyma.16G026300*, were retrieved in the soybean genome through sequence blast with AtCENH3 protein sequence, which is consistent with the previous study[28]. From the Phytozome database (https://phytozome-next.jgi.doe.gov/), four transcripts were predicted for *Glyma.07G057300*. However, through RT-PCR, only two of them, *Glyma.07G057300.1* and *Glyma.07G0573001.4*, were amplified in young inflorescences and young roots (Supplemental Fig. S1). The encoded Glyma.07G057300.1 protein is 49 amino acids longer than Glyma.07G057300.4 at the N-terminal end (Fig. 1a & b). Two transcripts were predicted for *Glyma.16G026300* on the Phytozome database (Supplemental Fig. S2), while only the shorter *Glyma.16G026300.2* version that encodes a truncated protein with an incomplete CATD (CENP-A target domain) (Fig. 1b), which is required for the loading of CENH3 onto the centromere[29−31], was detected. In addition, three unpredicted short transcripts, *Glyma.16G026300.3*, *Glyma.16G026300.4* and *Glyma.16G026300.5* were detected through RT-PCR and product sequencing (Supplemental Fig. S2).

To investigate the subcellular localization of the three identified transcripts of *CENH3*, we generated GFP reporter fusion constructs for Glyma.07G057300.1, Glyma.07G057300.4 and Glyma.16G026300.2, respectively. The constructs were expressed by transient transformation in soybean hairy roots. The result showed that GFP-Glyma.07G057300.1 and GFP-Glyma.07G057300.4 were localized on the centromeres shown by the dot GFP signal, while Glyma.16G026300.2 dispersed throughout the whole cell cytoplasm (Fig. 1c). This result suggests that *Glyma.07G057300* is the *bona fide CENH3* gene, and thus named *GmCENH3*.

Segregation distortion and aneuploidy occur in the offspring of null +/*Gmcenh3*

To investigate the function of *GmCENH3*, we used CRISPR/Cas9 gene-editing technology to mutate *GmCENH3* in the white-flowered variety Jack with sgRNAs targeting around the coding region of the glycine site right before the αN helix (Supplemental Fig. S3a), the point mutation of which site induced haploids in *Arabidopsis*[11,13]. We obtained five independent *GmCENH3*-edited lines: C10, C12, C16, C5 and C2. Among them, C10, C12 and C16 are biallelic with one allele of amino acid substitution or deletion within the open reading frame (ORF) and the other allele null mutation due to frame shift. C5 had two null alleles (+1 bp or +2 bp frameshift) and a wild-type (WT) allele, and C2 is biallelic with one allele of 4-amino-acid deletion and the other allele of four-amino-acid substitution (Fig. 2a & Supplemental Fig. S3b). Furthermore, a new biallelic null heterozygote C2-16 was obtained by crossing C2 and C16 (Fig. 2a). In the offspring of the null heterozygotes, we did not obtain homozygous seedlings of the null *Gmcenh3* alleles (Table 1), indicating lethality of homozygous embryo .

Notably, the segregation ratios of null heterozygotes C2-16 and C16, which lacked the glycine site, largely deviated from the Mendel Segregation law, while the segregation ratios of the null heterozygotes C12, C10, C5 and 4-aa-deletion/4-aa-substitution mutant C2 followed the law (Table 1). In C2-16, the 4-aa-deletion allele *cenh3-1* and null allele *cenh3-4* had comparable transmission efficiency in male gametes, while the transmission efficiency of *cenh3-4* in female gametes was too low to be detected (Table 2). In addition, among the progenies of the null heterozygotes in C2-16, C16 and C12, some null heterozygotes plants were totally sterile (Fig. 2b & c, 6/106 in C2-16, 5/86 in C16, 9/71 in C12). Genome re-sequencing of three of them, #609, #523, and #308, showed that these plants were aneuploids, with karyograms of 2n-1-1+1, 2n-1 and 2n-1, respectively (Fig. 2d). Unexpectedly, all these three aneuploids lost chromosome 6, suggesting biased chromosome elimination in the chromosome set. Therefore, knockout of *GmCENH3* causes genome instability in offspring of null heterozygotes.

Null *Gmcenh3* allele from the female gamete causes chromosome elimination

To examine whether the *GmCENH3*-edited lines (Jack) can induce haploids, we crossed them as the maternal to the white-flowered Wm82 and purple-flowered ZH13 varieties, respectively. Among the 164 F1 plants from crosses of null heterozygotes and Wm82, and the 711 F1 plants from crosses of amino-acid-change mutants and Wm82, we did not detect haploid plants using three pairs of molecular markers (Supplemental Table S1, Supplemental Fig. S4). It's worth noting that all these 164 F1 plants from crosses of null heterozygotes and Wm82 were fertile and inherited the amino acid change allele from the maternal (Table 2). In contrast, among the 18 F1 plants with purple flowers from crosses of null heterozygotes *cenh3-3*/*cenh3-4* and ZH13, we identified one sterile plant, #336, which had the genotype of +/*cenh3-4* and thus inherited the null *cenh3-4* allele from the maternal (Fig. 3a, Table 2). Genome re-sequencing of this plant showed that #336 lost chromosome 12 and chromosome 16 (Fig. 3b). Moreover, the two lost chromosomes were confirmed to be inherited from the maternal through 20 pairs of molecular makers between Jack and ZH13 that are distributed on the 20 nonhomologous chromosomes (Fig. 3c, Supplemental Fig. S5, Supplemental Table S2). The chromosomes in the null-allele-carrying female gamete were lost, indicating that the declined dosage of GmCENH3 impaired the function of centromeres in the female gametes and further led to the chromosome elimination.

Discussion

Our study identified the functional *GmCENH3* and validated its essential function in centromere assembly and seed production. Despite the whole genome duplication, loss of *CENH3* copies is common in many species, generating pseudogenes in the genome[32]. We confirmed that the truncated *Glyma.16G026300* does not produce a functional centromeric histone H3, rendering *Glyma.07G057300* as the sole functional *GmCENH3* in soybean. As reported previously[33], GmCENH3

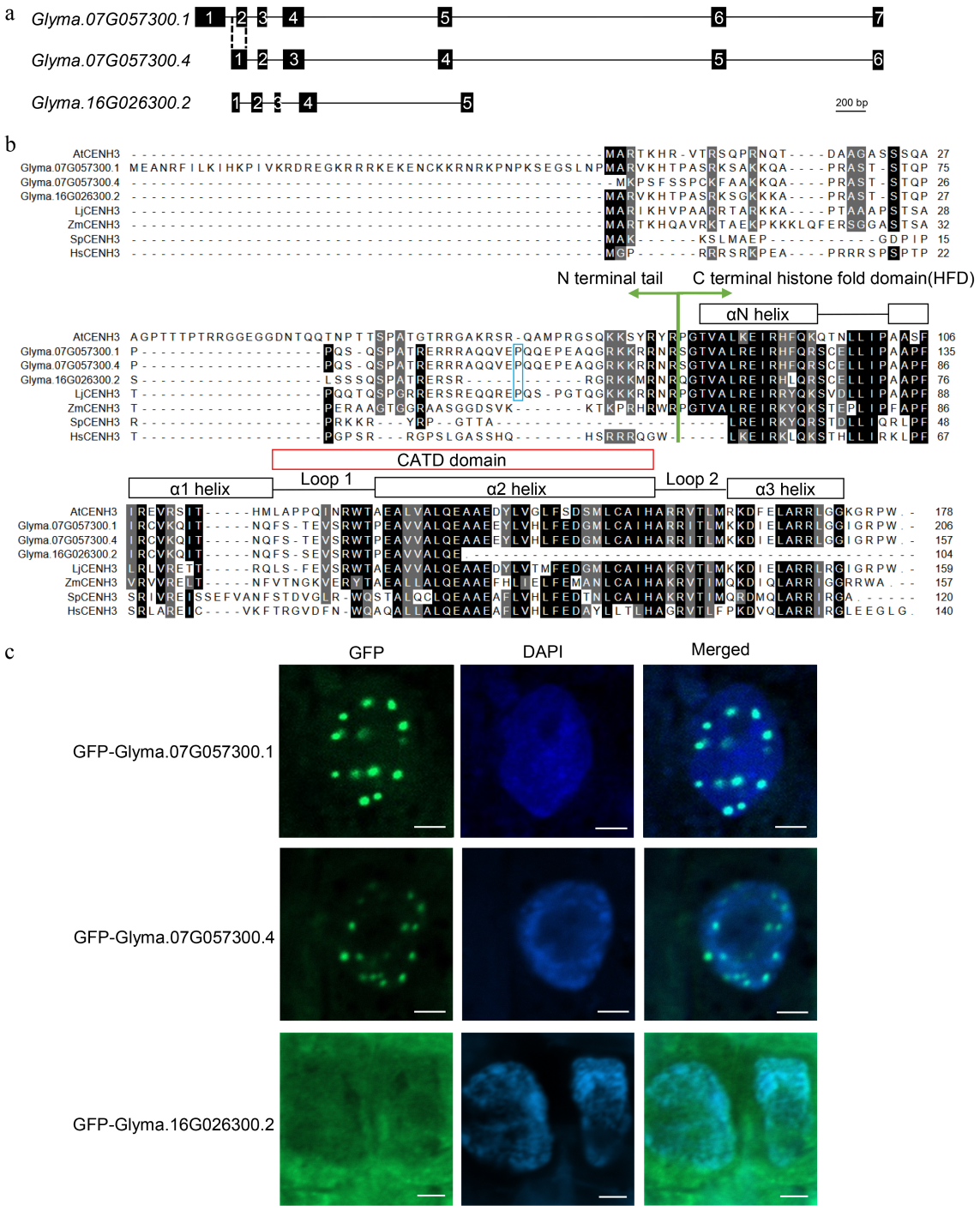

Fig. 1 Identification of CENH3 in soybean. (a) Transcript scheme of *Glyma.07G057300* and *Glyma.16G026300*. (b) Protein alignment of AtCENH3 (*Arabidopsis thaliana*, NP_001030927.1), Glyma.07G057300.1, Glyma.07G057300.4, Glyma.16G026300.2, LjCENH3 (*Lotus japonicus*, AHH01567.1), ZmCENH3 (*Zea mays*, NP_001105520.1), SpCENH3 (*Schizosaccharomyces pombe*, NP_596473.1) and HsCENH3 (*Homo sapiens*, NP_001800.1). (c) Subcellular locations of the GFP fusion proteins of Glyma.07G057300.1, Glyma.07G057300.4 and Glyma.16G026300.2. Scale bar = 2 μm.

proteins have a conserved C terminal histone fold domain and a variable N-terminal tail. GmCENH3.1 is 49 amino acids longer than GmCENH3.4 at the N-terminal tail and the two transcripts have similar expression levels in young inflorescences and young roots, raising the question that if the two GmCENH3 isoforms have different functions in binding centromeric DNA or other proteins through epigenetic modification and if they can be assembled into the same nucleosome.

Through gene editing and outcrossing, we obtained five null heterozygous lines of *GmCENH3* with homozygote-lethal phenotype. In the two null heterozygous lines C2-16 and C16, the glycine site, which was reported to induce haploid when mutated in *Arabidopsis*[11,13], was missing in the amino-acid-deletion allele, and coincidently, segregation distortion occurred only in these two lines rather than the other three null heterozygote lines (C12, C10 and C5) without this site mutation

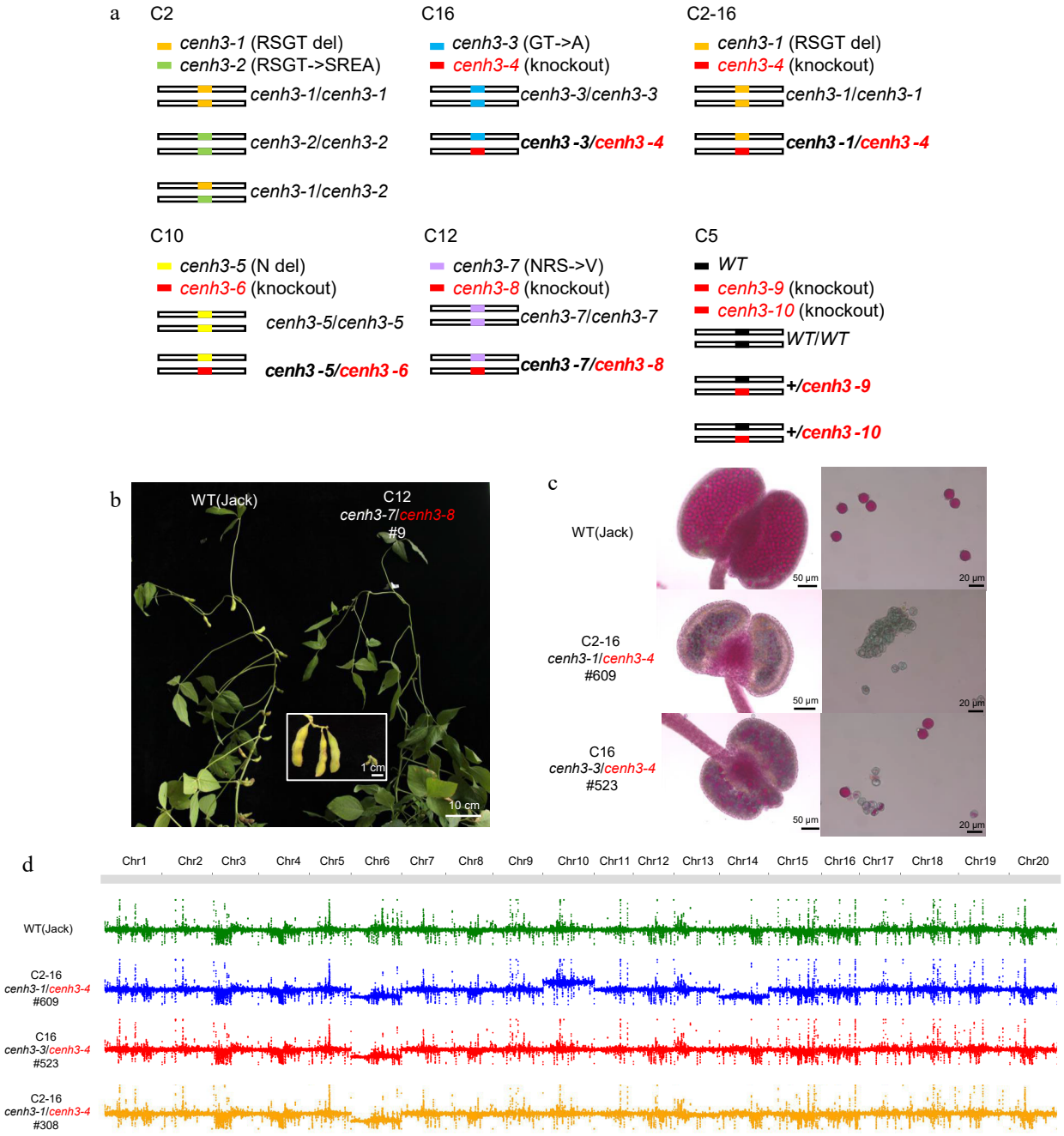

Fig. 2 Heterozygosity of *GmCENH3* causes genome instability. (a) Diagram of genotypes of six lines with edited *GmCENH3*. Null alleles were marked by red font and the null heterozygotes were marked by bold font. (b) A plant with the genotype *cenh3-7/cenh3-8* among the progeny of null heterozygote C12 was totally sterile. Scale bar = 10 cm. (c) The inviable pollen in the two sterile plants (#609 and #523) among the progeny of null heterozygotes C2-16 and C16. Pollen of plant #609 with genotype *cenh3-1/cenh3-4* were totally inviable, and pollen of plant #523 with genotype *cenh3-3/cenh3-4* were partially inviable. Pollen were stained by Alexander staining. (d) #609, #523 and #308 had abnormal chromosome numbers.

(Fig. 2, Table 1). In maize, more severe segregation distortion occurred in +/*cenh3* null heterozygotes and the transmission efficiency of null allele was much lower than WT in both male and female gamete[15]. By contrast, in *Arabidopsis* the segregation ratio of null heterozygote +/*Atcenh3* is in agreement with Mendel's law of segregation[34]. In addition, considering that the null allele *Gmcenh3-4* and the 4-aa-deletion allele *Gmcenh3-1* in C2-16 line had comparable transmission efficiency in the male gametes (Table 2). It is possible that the egg cells with null *Gmcenh3* allele and residual glycine-deleted Gmcenh3 proteins had lower transmission efficiency than the egg cells with null alleles and residual glycine-reserved Gmcenh3 proteins in C12, C10 and C5 lines. However, distinct from the high haploid induction rates (HIR) of similar mutants in *Arabidopsis*[11,13], the *Gmcenh3-1*, *Gmcenh3-2* or *Gmcenh3-3* alleles in C2, C16 and C2-16 lines expressing glycine-mutated Gmcenh3 protein did

Table 1. Progeny segregation of edited-*GmCENH3*.

Genotype	Number of plants	Expected number	Chi-square test
C2 (*cenh3-1*/*cenh3-2*)			
cenh3-1/*cenh3-1*	17	15	
cenh3-2/*cenh3-2*	15	15	0.9086
cenh3-1/*cenh3-2*	29	31	
C16 (*cenh3-3*/*cenh3-4*)			
cenh3-3/*cenh3-4*	102	133	0.0016**
cenh3-3/*cenh3-3*	97	66	
C2-16 (*cenh3-1*/*cenh3-4*)F2			
cenh3-1/*cenh3-4*	21	123	<0.0001****
cenh3-1/*cenh3-1*	163	61	
C2-16 (*cenh3-1*/*cenh3-4*)F3, F4			
cenh3-1/*cenh3-4*	130	174	<0.0001****
cenh3-1/*cenh3-1*	131	87	
C10 (*cenh3-5*/*cenh3-6*)			
cenh3-5/*cenh3-6*	40	46	0.2919
cenh3-5/*cenh3-5*	29	23	
C12 (*cenh3-7*/*cenh3-8*)			
cenh3-7/*cenh3-8*	102	107	0.5570
cenh3-7/*cenh3-7*	58	53	
C5-4 (+/*cenh3-9*)			
+/*cenh3-9*	45	50	0.3969
+/+	30	25	
C5-40 (+/*cenh3-10*)			
+/*cenh3-10*	55	63	0.2316
+/+	40	32	

Red: null alleles.

not induce chromosome elimination in soybean (Supplemental Table S1).

Deposition of new CENH3 is limited to the existing CENH3 domain to avoid abnormal neocentromere formation, thus maintaining enough dosage of CENH3 on centromere is important for CENH3 licensing and centromere function[32,35]. In null heterozygous zygote, the deposition of newly synthesized uniparental CENH3 onto the centromeres in the null-allele-carrying gamete may be hindered, leading to failure of kinetochore assembly and chromosome elimination. This hypothesis was supported by the haploid production resulted from null heterozygotes in *Arabidopsis* and maize and biased CENH3 loading in early *cenh3-1; GFP-ts/CENH3* hybrid embryos in *Arabidopsis*[14,15]. The HIR of null alleles from egg cell were 0.83% and 10% in *Arabidopsis* and maize, respectively. Since AtCENH3 proteins persist on the egg cell centromeres[14], a functional megaspore with null allele genotype could not produce an egg cell with fully functional centromeres after three round of mitoses with little residual CENH3 on centromeres, resulting occasionally maternal chromosome elimination and haploid induction. In contrast, the transmission efficiency of null alleles from soybean female gametes with residual glycine-deleted Gmcenh3 proteins was extremely low, but once it passes down, it induces chromosome elimination and produced aneuploids (Table 2, Supplemental Table S1). Loss-of-CENH3 induced aneuploids and haploids are both caused by chromosome elimination, and their difference is the number of lost chromosomes[9]. Hence, it would be promising to induce haploid through designed gene-editing of *GmCENH3* to achieve a balance between female transmission, CENH3 turnover velocity and chromosome elimination.

There are several possible reasons for the lack of haploid in the progeny of the crossed null heterozygous C2-16 or C16. Firstly, the functional megaspore with null allele in C2-16 or C16 produces malfunctional female gametes with extremely weak centromeres carrying residual glycine-lacking cenh3 proteins. This scenario is supported by the low transmission efficiencies of the null *cenh3* through the female gametes in C2-16 and C16 (Table 2) and the HI effect of the glycine lacking in *Arabidopsis*[11,13]. Thus, the functional megaspore with null allele in C12, C10 and C5 with normal transmission efficiency could produce functional female gametes with stronger centromeres carrying residual glycine-reserved cenh3 proteins and then the female gametes with null alleles pass down to induce aneuploids and even haploids. Another scenario is that the null allele causes incomplete and asynchronous chromosome eliminations due to the large numbers of chromosome in soybean (2n = 40) and consequently embryo lethality. And according to the fact that the survived aneuploids had eliminations of only 1 to 3 chromosomes (Figs 2 & 3), soybean may be less tolerant to genome instability, causing aneuploid or haploid embryos underdeveloped. It's also possible that the newly expressed paternal WT GmCENH3 proteins rapidly recover the impaired female centromeres before large scale chromosome elimination. Genetic engineering of *GmCENH3* to produce an N-terminal tail changed protein might be more efficient than the null allele in the competition with WT GmCENH3[10]. The last scenario taken into consideration is the different genetic background of the parental varieties used in outcrossing. In the cross between null heterozygotes and ZH13, the transmission efficiency of null alleles through female gametes is 5.6%, but the transmission efficiency was zero in the massive crosses between null heterozygotes and Wm82. To confirm whether the protein sequence difference of GmCENH3 among Jack, Wm82 and ZH13 caused the contrasting result, we sequenced the CDS of *GmCENH3* from the three varieties and found that the 284th cytosine of Glyma.07G057300 CDS1 in ZH13, changed to adenine, leading to a proline-to-glutamine change at the N-terminal tail (Fig. 1b, Supplemental Fig. S1). Considering the difference of centromeric DNA sequences among soybean accessions[36], the difference of centromeres between Jack, Wm82 and ZH13 may contributes to the distinct transmission efficiency, which need further experimentation.

Table 2. Transmission efficiencies of *cenh3-1*, *cenh3-3* and *cenh3-4*.

Crosses	Number of F1 plants	Number of +/*cenh3-x* (x = 1,3)	Number of +/*cenh3-4*
cenh3-1/*cenh3-4*(Jack) ♀ × WT(Wm82) ♂	94	94	0
cenh3-1/*cenh3-4*(Jack) ♀ × WT(Jack) ♂	16	16	0
WT(Jack) ♀× *cenh3-1*/*cenh3-4*(Jack) ♂	28	13	15
cenh3-3/*cenh3-4*(Jack) ♀ × WT(Wm82) ♂	70	70 (11 aborted seeds)	0
cenh3-3/*cenh3-4*(Jack) ♀ × WT(ZH13) ♂	18	17 (1 aborted seed)	1

Red: null alleles.

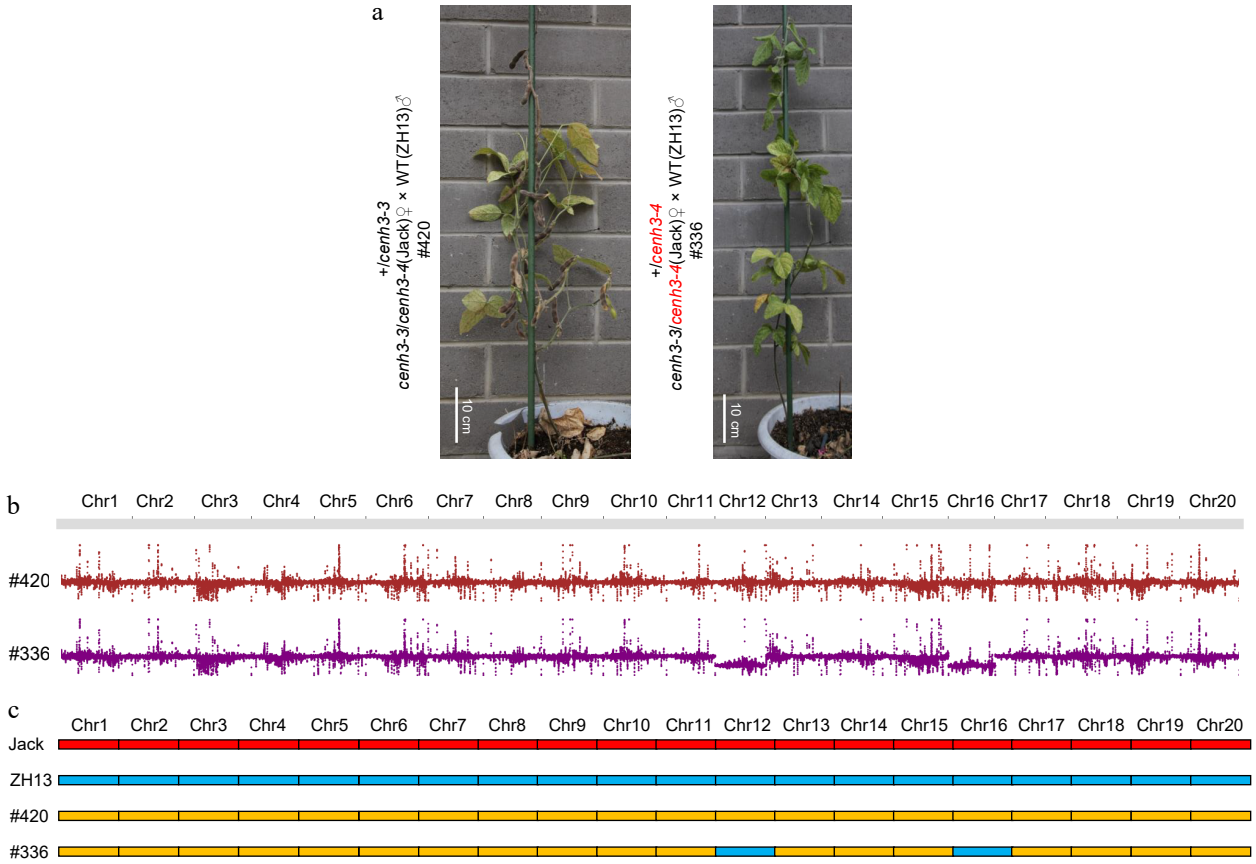

Fig. 3 Female gamete of null allele causes chromosome elimination. (a) 5-month-old F1 plants from the parental null heterozygotes *cenh3-3/cenh3-4* and ZH13. #420 was fertile while #336 was sterile. Scale bar = 10 cm. (b) #420 had 40 chromosomes while #336 lost one chromosome 12 and one chromosome 16. (c) The two lost chromosomes in #336 originated from the null heterozygote *cenh3-3/cenh3-4*. Homozygous molecular markers from Jack were marked as red, and homozygous molecular markers from ZH13 were marked as blue, and heterozygous molecular markers between Jack and ZH13 were marked as yellow.

Author contributions

The authors confirm contribution to the paper as follows: research design: Li HJ, Yang WC; experiment performance: Wang J; bioinformatic analysis: Wang XF; draft manuscript preparation: Wang J, Li HJ. All authors reviewed the results and approved the final version of the manuscript.

Data availability

All data generated or analyzed during this study are included in this published article and its Supplementary information files.

Acknowledgments

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Science (XDA24020306), National Natural Science Foundation of China (31991203), the National Key Research and Development Program of China (2022YFF1003500), CAS Project for Young Scientists in Basic Research (No.YSBR-078), and Hainan Yazhou Bay Seed Laboratory and China National Seed Group (project of B23YQ1505). We thank Prof. Zhixi Tian (Institute of Genetics and Developmental Biology, CAS) for vectors used in gene-editing, ZH13 seeds and instruction in soybean crossing. We thank Beijing Genomics Institute for their support in genome re-sequencing and bioinformatic analysis. We also thank Jie Luo and Wenqian Zhang for help in soybean crossing.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary Information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/SeedBio-2023-0024)

Dates

Received 10 October 2023; Accepted 5 December 2023; Published online 26 December 2023

References

1. McKinley KL, Cheeseman IM. 2016. The molecular basis for centromere identity and function. *Nature Reviews Molecular Cell Biology* 17:16–29
2. Comai L, Maheshwari S, Marimuthu MPA. 2017. Plant centromeres. *Current Opinion in Plant Biology* 36:158–67
3. Black BE, Jansen LET, Maddox PS, Foltz DR, Desai AB, et al. 2007.

Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Molecular Cell* 25:309−22
4. Musacchio A, Desai A. 2017. A molecular view of kinetochore assembly and function. *Biology* 6:5
5. Fukagawa T, Earnshaw WC. 2014. The Centromere: Chromatin Foundation for the Kinetochore Machinery. *Developmental Cell* 30: 496−508
6. Guo W, Comai L, Henry IM. 2023. Chromoanagenesis in plants: triggers, mechanisms, and potential impact. *Trends in Genetics* 39: 34−45
7. Ravi M, Chan SWL. 2010. Haploid plants produced by centromere-mediated genome elimination. *Nature* 464:615−18
8. Kelliher T, Starr D, Wang W, McCuiston J, Zhong H, et al. 2016. Maternal haploids are preferentially induced by *CENH3-tailswap* transgenic complementation in maize. *Frontiers in Plant Science* 7: 414
9. Maheshwari S, Tan EH, West A, Franklin FC, Comai L, et al. 2015. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genetics* 11: e1004970
10. Lv J, Yu K, Wei J, Gui H, Liu C, et al. 2020. Generation of paternal haploids in wheat by genome editing of the centromeric histone *CENH3*. *Nature Biotechnology*1397−401
11. Kuppu S, Tan EH, Nguyen H, Rodgers A, Comai L, et al. 2015. Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. *PLoS Genetics* 11:e1005494
12. Karimi-Ashtiyani R, Ishii T, Niessen M, Stein N, Heckmann S, et al. 2015. Point mutation impairs centromeric CENH3 loading and induces haploid plants. *Proceedings of the National Academy of Sciences of the United States of America* 112:11211−16
13. Kuppu S, Ron M, Marimuthu MPA, Li G, Huddleson A, et al. 2020. A variety of changes, including CRISPR/Cas9-mediated deletions, in CENH3 lead to haploid induction on outcrossing. *Plant Biotechnology Journal* 18:2068−80
14. Marimuthu MPA, Maruthachalam R, Bondada R, Kuppu S, Tan EH, et al. 2021. Epigenetically mismatched parental centromeres trigger genome elimination in hybrids. *Science Advances* 7:eabk1151
15. Wang N, Gent JI, Dawe RK. 2021. Haploid induction by a maize cenh3 null mutant. *Science Advances* 7:eabe2299
16. Demidov D, Lermontova I, Moebes M, Kochevenko A, Fuchs J, et al. 2022. Haploid induction by nanobody-targeted ubiquitin-proteasome-based degradation of EYFP-tagged CENH3 in *Arabidopsis thaliana*. *Journal of Experimental Botany* 73:7243−54
17. Li S, Wang J, Jia S, Wang K, Li H. 2023. Synthetic apomixis: from genetic basis to agricultural application. *Seed Biology* 2:10
18. Sanei M, Pickering R, Kumke K, Nasuda S, Houben A. 2011. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proceedings of the National Academy of Sciences of the United States of America* 108:E498−E505
19. Kenworthy WJ, Brim CA, Wernsman EA. 1973. Polyembryony in soybeans. *Crop Science* 13:637−39
20. Clarke JD. 2009. Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harbor protocols* 2009:pdb.prot5177
21. Concordet JP, Haeussler M. 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research* 46:W242−W245
22. Sun B, Wang Y, Yang Q, Gao H, Niu H, et al. 2023. A high-resolution transcriptomic atlas depicting nitrogen fixation and nodule development in soybean. *Journal of Integrative Plant Biology* 65:1536−52
23. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17:10−12
24. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754−60
25. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008
26. Van der Auwera GA & O'Connor BD. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition).
27. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6:80−92
28. Ishii T, Juranić M, Maheshwari S, de Oliveira Bustamante F, Vogt M, et al. 2020. Unequal contribution of two paralogous CENH3 variants in cowpea centromere function. *Communications Biology* 3:775
29. Bassett EA, DeNizio J, Barnhart-Dailey MC, Panchenko T, Sekulic N, et al. 2012. HJURP Uses Distinct CENP-A Surfaces to Recognize and to Stabilize CENP-A/Histone H4 for Centromere Assembly. *Developmental Cell* 22:749−62
30. Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL, et al. 2004. Structural determinants for generating centromeric chromatin. *Nature* 430:578−82
31. Pan D, Walstein K, Take A, Bier D, Kaiser N, et al. 2019. Mechanism of centromere recruitment of the CENP-A chaperone HJURP and its implications for centromere licensing. *Nature Communications* 10:4046
32. Neumann P, Pavlíková Z, Koblížková A, Fuková I, Jedličková V, et al. 2015. Centromeres off the hook: Massive changes in centromere size and structure following duplication of *CenH3* gene in *Fabeae* species. *Molecular Biology and Evolution* 32:1862−79
33. De Rop V, Padeganeh A, Maddox PS. 2012. CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly. *Chromosoma* 121:527−38
34. Ravi M, Kwong PN, Menorca RMG, Valencia JT, Ramahi JS, et al. 2010. The rapidly evolving centromere-specific histone has stringent functional requirements in *Arabidopsis thaliana*. *Genetics* 186: 461−71
35. Hoffmann S, Izquierdo HM, Gamba R, Chardon F, Dumont M, et al. 2020. A genetic memory initiates the epigenetic loop necessary to preserve centromere position. *The EMBO Journal* 39:e105505
36. Liu Y, Yi C, Fan C, Liu Q, Liu S, et al. 2023. Pan-centromere reveals widespread centromere repositioning of soybean genomes. *Proceedings of the National Academy of Sciences of the United States of America* 120:e2310177120

Copyright: © 2023 by the author(s). Published by Maximum Academic Press on behalf of Hainan Yazhou Bay Seed Laboratory. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0/.