



# Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3

Jian Lv<sup>1,4</sup>✉, Kun Yu<sup>1,4</sup>, Juan Wei<sup>1</sup>, Huaping Gui<sup>1</sup>, Chunxia Liu<sup>1</sup>, Dawei Liang<sup>1</sup>, Yanli Wang<sup>1</sup>, Hongju Zhou<sup>1</sup>, Ryan Carlin<sup>2</sup>, Randy Rich<sup>3</sup>, Tiancong Lu<sup>1</sup>, Qiudeng Que<sup>2</sup>, Wen Chung Wang<sup>1</sup>, Xingping Zhang<sup>1</sup> and Tim Kelliher<sup>1,2</sup>✉

**New breeding technologies accelerate germplasm improvement and reduce the cost of goods in seed production<sup>1–3</sup>. Many such technologies could use *in vivo* paternal haploid induction (HI), which occurs when double fertilization precedes maternal (egg cell) genome loss. Engineering of the essential CENTROMERIC HISTONE (CENH3) gene induces paternal HI in *Arabidopsis*<sup>4–6</sup>. Despite conservation of CENH3 function across crops, CENH3-based HI has not been successful outside of the *Arabidopsis* model system<sup>7</sup>. Here we report a commercially operable paternal HI line in wheat with a ~7% HI rate, identified by screening genome-edited *TaCENH3α*-heteroallelic combinations. Unlike in *Arabidopsis*, edited alleles exhibited reduced transmission in female gametophytes, and heterozygous genotypes triggered higher HI rates than homozygous combinations. These developments might pave the way for the deployment of CENH3 HI technology in diverse crops.**

Doubled haploid (DH) technology accelerates genetic gain by enabling the rapid production of 100% homozygous inbred lines<sup>1,2</sup>. HI might be achieved *in vivo* (in seeds) or *in vitro* by gametophyte culture. *In vivo* HI relies on mutant alleles triggering aberrant reproductive states, such as the frameshift mutation in a pollen-specific phospholipase MATRILINEAL (MATL; also known as NOT LIKE DAD (NLD) and PHOSPHOLIPASE A1 (PLA1))<sup>8–10</sup>. Knockout of MATL/NLD/PLA1 orthologs in rice<sup>11</sup> or wheat<sup>12,13</sup> leads to HI. Current reports indicate that MATL is grass specific<sup>11</sup>, although similar DMP mutants induce a moderate-rate HI in *Arabidopsis*<sup>14</sup>. Meanwhile, new applications of HI systems have been reported, including HI-Edit<sup>3</sup>, which introduces genome-editing components to elite varieties via an HI line. This process produces transgene-free edited DH lines.

One promising approach to HI engineering is the CENTROMERIC HISTONE3 (CENH3) gene, comprising a highly conserved C-terminal histone fold domain (HFD) and a hypervariable N-terminal tail that protrudes from the nucleosome core particle (Fig. 1d). The tail is involved in the recruitment and stabilization of centromere complex proteins, a feature that was exploited to give about 30% intraspecific paternal HI<sup>4</sup> in *Arabidopsis* after transgenic replacement with the tail of histone H3.3 fused to green fluorescent protein (GFP). This vector partially complemented *cenh3* null mutants but triggered male sterility. On outcrossing, GFP-tagged chromosomes deriving from the egg cell were lost in embryogenesis, generating roughly equal tallies of haploid, diploid and aneuploid offspring. In subsequent work, amino acid substitutions and in-frame deletions in *AtCENH3* induced haploids as well<sup>5,6</sup>, but

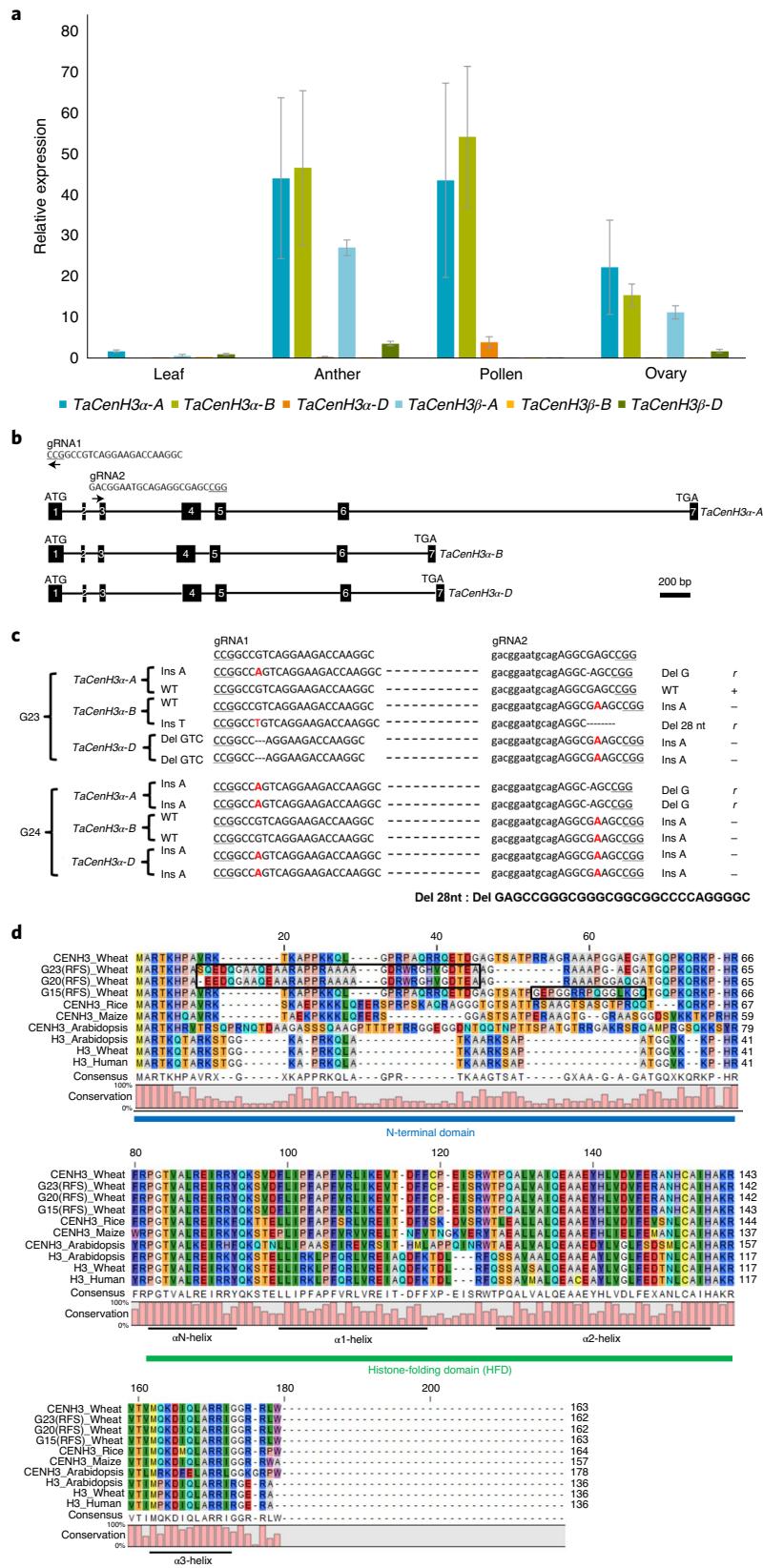
evaluation of a few of the same point mutations in rice *CENH3* did not induce haploids (unpublished data, C.M. and Y.P.). Although *CENH3* function is conserved across kingdoms, an efficient HI effect has not been induced outside of *Arabidopsis*<sup>15</sup>. One study in maize achieved a 3.6% haploid induction rate (HIR) in one cross, but the average HIR for all events was less than 1%<sup>7</sup>. In this study, we used genome editing to create restored frameshift (RFS, or *r*) tail-modified alleles of wheat *CENH3α* while leaving the canonical HFD domain intact. The guide RNAs (gRNAs) that created RFS alleles also knocked out other homoeologs, which could help the RFS alleles become predominant protein variants.

Bread wheat is a hexaploid crop comprising three subgenomes, named A, B and D. Wheat and barley have two versions of *CENH3* ( $\alpha$  and  $\beta$ ), which localize to complementary centromeric domains<sup>16</sup>. Silencing of *TaCENH3α* via RNA interference (RNAi) leads to dwarfed plants; RNAi to *TaCENH3β* reduces plant height and reproductive fitness<sup>16</sup>. We analyzed the expression of the A, B and D genome's homoeologs of *CENH3α* and *CENH3β* from the spring wheat variety Fielder. *TaCenH3α* expression was higher than *TaCenH3β* in ovaries and pollen (Fig. 1a); homoeolog-specific quantitative real-time polymerase chain reaction (qRT-PCR) clarified that *TaCenH3α-A* and -B were highly expressed; and *TaCenH3α-D* was below the detection limit (Fig. 1a).

To create knockout and hypomorphic alleles, we designed CRISPR-Cas9 constructs containing two single gRNAs (gRNA1 and gRNA2 in construct 24195; gRNA3 and gRNA4 in construct 24194) with protospacer sequences matching the three homoeologs of *TaCENH3α* but with more than 12 mismatches to *TaCENH3β* and no predicted off-targets (Fig. 1b). Whereas gRNA1 targets the start of the N-terminal domain in exon 1, gRNA2 targets the intron 2/exon 3 splice site (Fig. 1b). We sequenced *CENH3α* alleles from 1,152 TaqMan-positive events. No plants in the T<sub>0</sub> or T<sub>1</sub> generations had all six *CENH3α* alleles knocked out, nor were any recovered with both copies of the -A homoeolog knocked out. These results suggest that *CENH3α-A* is required for plant development or reproduction.

We grew 702 T<sub>0</sub> events to maturity, focusing especially on those with edits in all homoeologs and restored frameshift (*r*) alleles. One such T<sub>0</sub> plant of interest had the genotype (+/*r*, -/*r*, -/-): that is, heterozygous for -A with one wild-type and one RFS allele; heterozygous for -B with one knockout and one RFS allele; and homozygous knockout for -D. The RFS in *TaCENH3α-A* derived from an adenine insertion at gRNA1 and a deletion of guanine at gRNA2 (Fig. 1c). The first mutation shifted the open reading frame, but

<sup>1</sup>Seeds Research, Syngenta Biotechnology China, ZhongGuanCun Life Science Park, Beijing, China. <sup>2</sup>Seeds Research, Syngenta Crop Protection, LLC, Research Triangle Park, NC, USA. <sup>3</sup>Seeds Research, Syngenta Seeds Research, Junction City, KS, USA. <sup>4</sup>These authors contributed equally: Jian Lv, Kun Yu.  
✉e-mail: [jian.lv@syngenta.com](mailto:jian.lv@syngenta.com); [tim.kelliher@syngenta.com](mailto:tim.kelliher@syngenta.com)



**Fig. 1 | Genome editing strategy and generation of HI alleles.** **a**, The expression of the *CENH3 $\alpha$*  and *CENH3 $\beta$*  A, B and D homoeologues in leaf and reproductive tissues indicates that *CENH3 $\alpha$*  predominates in sexual structures. The low expression in the elongated leaf likely reflects less active cell division there. **b**, The locations of gRNA1 and gRNA2 are shown over the exon/intron structure of *CENH3 $\alpha$* . **c**, *CENH3 $\alpha$*  A, B and D homoeologues edited allele genotypes in the G23 and G24 HI plants. **d**, Amino acid alignment of *TaCENH3 $\alpha$ -A* and *r* alleles compared to *CENH3* and *H3* in various species. N-terminal hypervariable and C-terminal HFDs are underlined in blue and green bars. Detailed RFS changes in the G15 and G23/G24 genotypes are outlined in black boxes.

**Table 1 | Outcrossing paternal HIR data of different genotypic classes in T<sub>1</sub> and T<sub>2</sub> generations**

Class	Construct	Generation	Genotype	RFS or knockout alleles	No. F <sub>1</sub> plants	No. haploids	Outcross HIR
G15	24194	T <sub>1</sub>	r/r, -/-, Δ/Δ	A: ΔG at gRNA3, ^A at gRNA4 B: ^T at gRNA4 D: Δ648 bp between gRNAs	208	1	0.5%
G18	24195	T <sub>1</sub>	+/, -, -, -/-	A: WT at gRNA1 and gRNA2 B: ^A at gRNA2 D: ^T at gRNA2	150	0	0
G19	24195	T <sub>1</sub>	+/-, Δ/Δ, Δ/Δ	A: ^T at gRNA2 B: Δ332 bp at gRNA1 D: Δ24 bp at gRNA1	196	0	0
G20	24195	T <sub>1</sub>	r/r, -/-, Δ/Δ	A: ΔTC at gRNA1, ΔG at gRNA2 B: ^T at gRNA2	176	0	0
G21	24195	T <sub>1</sub>	r/r, Δ/Δ, Δ/Δ	A: ΔTC at gRNA1, ΔG at gRNA2 B: Δ332 bp at gRNA1 D: Δ24 bp at gRNA1	168	0	0
G22	24195	T <sub>1</sub>	+/-, r/r, -/-	A: ^T at gRNA1 B: ^T at gRNA1, Δ28nt at gRNA2 D: ΔCGT at gRNA1, ^A at gRNA2	170	0	0
G23	24195	T <sub>1</sub>	+/r, -/-, -/-	A: ^A at gRNA1, ΔG at gRNA2 B: ^A at gRNA2 D: ΔGTC at gRNA1, ^A at gRNA2	57	4	7.0%
G24	24195	T <sub>1</sub>	r/r, -/-, -/-	A: same as G23; B: ^A at gRNA2; D: ^A at gRNA1, ^A at gRNA2	143	1	0.7%
G23	24195	T <sub>2</sub>	+/r, -/-, -/-	Same as G23-T <sub>1</sub>	50	4	8.0%
G24	24195	T <sub>2</sub>	r/r, -/-, -/-	Same as G24-T <sub>1</sub>	55	1	1.8%

The number of haploids is determined by KASP marker assay and confirmed by ploidy analysis. Note: (r) refers to an RFS allele; (Δ) refers to a predicted in-frame deletion; and (+) represents a wild-type (unedited) allele, whereas (−) refers to a predicted frameshift (knockout).

the complementary guanine insertion 34 amino acids downstream restored it back to wild type. The new RFS sequence alters putative histone modification sites (Fig. 1d).

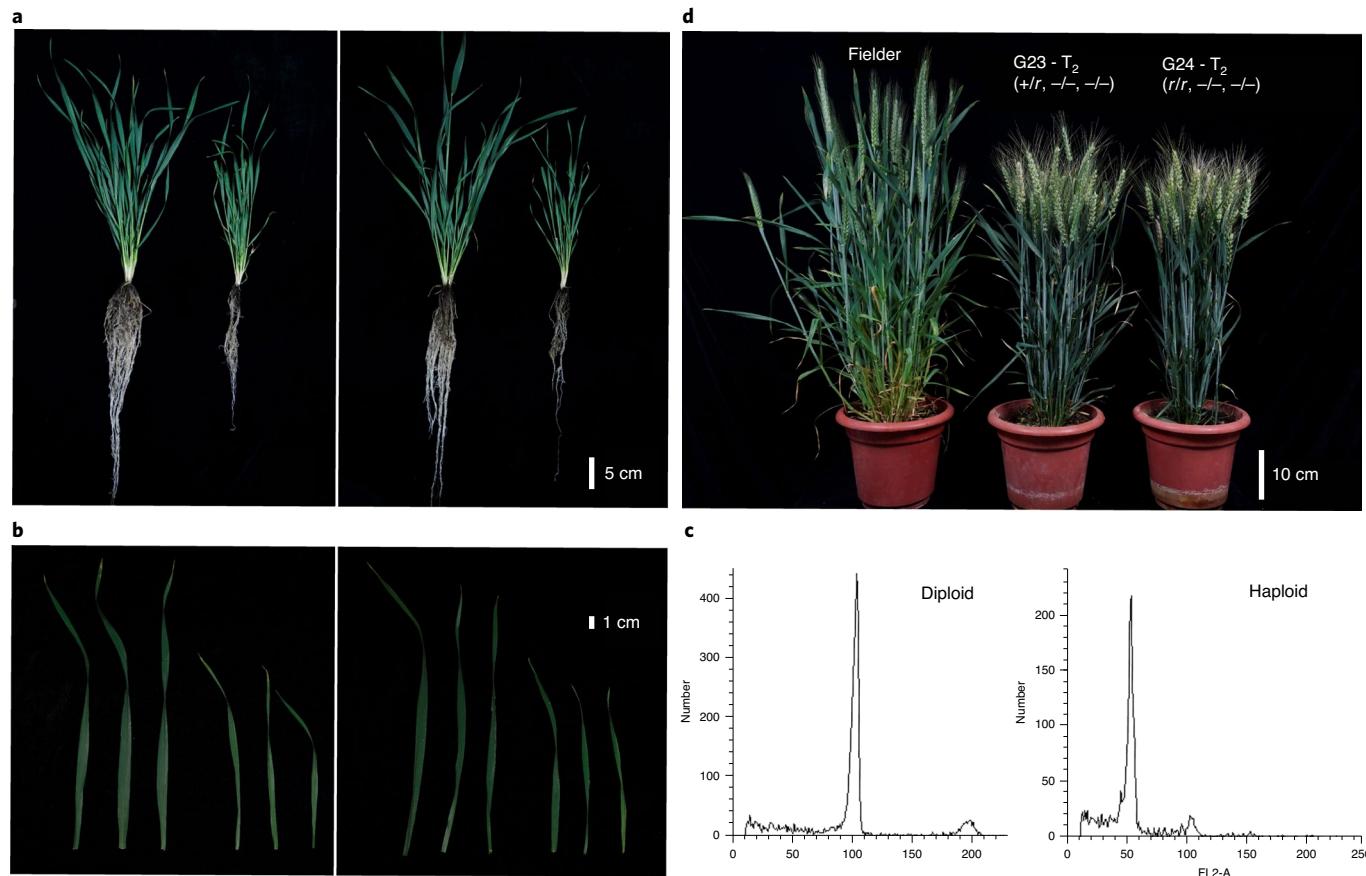
Selected T<sub>1</sub> genotypes from several T<sub>0</sub> events of interest were grown to maturity, allowed to self-pollinate and back-crossed as females to the wild-type line 03S0352-22 to test the outcrossing HIR (Supplementary Fig. 1). In normally germinated seedlings, putative haploid progeny were identified using KASP assays (Supplementary Table 1) and confirmed by flow cytometry. T<sub>1</sub> progeny of the T<sub>0</sub> event mentioned above with a 'G23' genotype (+/r, -/-, -/-) (Fig. 1c) led to 6.9% (9/131) haploids by self-pollination, and, after outcrossing, the paternal HIR was 7.0% (4/57) (Table 1, Fig. 2a–c and Supplementary Table 2). T<sub>1</sub> plants without a hypomorphic RFS allele in the A genome, such as G19 (+/-, Δ/Δ, Δ/Δ where Δ signifies an in-frame deletion), G18 (+/, -, -, -/-) and G22 (+/-, r/r, -/-), did not induce haploids (Table 1). These results are consistent with the hypothesis that an RFS (r) allele in CENH3α-A triggers HI.

The T<sub>1</sub> plants used for HIR testing contained the editing transgene. Genotyping seedling stage leaves did not detect any further editing of CENH3, but, to confirm that the G23 genotype (+/r, -/-, -/-) produces a high HIR, T<sub>2</sub> G23 plants were tested again, and the outcrossing paternal HIR was 8% (4/50) (Table 1). This confirms that the heterozygous (+/r) combination of an RFS allele and an unmodified TaCENH3α-A allele plus knockout of both copies of the -B and -D homoeologs led to the high HIR. Reasoning that homozygous lines produce uniform gametophyte genotypes and, thus, might have higher HIRs than heterozygous inducers, we tested the HIR of 'G24' T<sub>1</sub> plants with the genotypic combination (r/r, -/-, -/-) (Fig. 1c). Surprisingly, these plants produced only 0.7% and 1.8% outcrossing HIR in T<sub>1</sub> and T<sub>2</sub> generations, respectively (Table 1). Taken together, these results suggest that high HIRs

in wheat derive from a +/r heterozygous state for the A homoeolog of CENH3α. Haploids generated in this study were smaller than diploids (Fig. 2a,b), reminiscent of maternal haploids induced by knockout of TaMATEL<sup>12,13</sup>. Haploid leaf stomata were smaller than those of diploids (Supplementary Fig. 2).

To test whether other N-terminal RFS alleles can trigger HI, we analyzed the T<sub>1</sub> events from construct 24194, which had two different gRNAs (gRNA3 and gRNA4; Supplementary Fig. 1). The T<sub>1</sub> G15 genotype (r/r, -/-, Δ/Δ) had a 0.5% paternal induction rate (Table 1; 1/208 F<sub>1</sub> plants), mirroring the HIR in the G24 homozygous combination (r/r, -/-, -/-). The TaCenH3α-A-restored frameshift (r) allele in G15 had a different sequence than that in G23 and G24 (Fig. 1d).

G23 and G24 inducers were checked for pleiotropic vegetative and reproductive phenotypes. Besides being slightly shorter than wild-type plants (Fig. 2d), these inducers flowered earlier, and spike length was reduced (Supplementary Fig. 3). Together, it seems that the RFS alleles that enable HI might also accelerate the life cycle. Gross floral morphology was normal; no difference was found between pollen quality as measured by iodine staining (Supplementary Fig. 3), but the reproductive transmission of RFS alleles was compromised. Instead of a 1:2:1 segregation pattern of the G23 CENH3α-A (r) allele, we found few (r/r) and (+/r) progeny (Supplementary Table 3). In outcrosses to 03S0352-22 pollen, again we found segregation distortion (Supplementary Tables 4 and 5), suggesting that allele transmission is reduced through the female gametophyte. The G23 self-pollinated seed set was just 46.8%, but the seed set of G24 plants was normal (Supplementary Table 6). Intriguingly, these data are consistent with the possibility that the heterozygous (+/r) combination weakens female gamete performance but enhances HI.



**Fig. 2 | Haplid and inducer phenotypes.** **a**, Diploid  $F_1$  (left) and haploid (right) offspring from an outcross of the G23 inducer (left panel) and the G24 inducer (right panel) with the male pollen donor O3S0352-22. **b**, Diploid  $F_1$  (left) and haploid (right) leaves from G23 (left panel) and G24 (right panel) inducers. **c**, Ploidy analysis results from a representative G23 haploid and an  $F_1$  diploid control. **d**, Diploid, unedited Fielder wheat and G23 and G24 genome-edited plants showing altered vegetative characteristics.

Another common feature of *Arabidopsis* *CENH3*-tailswap inducer lines was the large number of aneuploid progenies produced via incomplete female genome elimination<sup>4</sup>. We did not check aneuploid plants in G23 or G24 outcrosses, but 23 (17.6%) putative aneuploids were identified among 131  $T_2$  offspring of G23 (+/r, -/-, -/-) (Supplementary Fig. 4).

The 2010 report that engineering of the N-terminal domain of *CENH3* triggers highly efficient HI in *Arabidopsis*<sup>4</sup> supported the potential of engineering DH systems in crops<sup>15</sup>. Since then, an array of point mutations<sup>5,6,17</sup>, in-frame deletions<sup>5</sup> and whole-gene replacements from other species<sup>18</sup> have triggered 0.6–44%<sup>5</sup> HI in *Arabidopsis*. However, no *CENH3*-based HI systems are currently used in breeding programs, mostly owing to the low HIR found in crop species<sup>7,15</sup>. Recent patents claim that single amino acid substitutions generated HI in cucurbits and other dicots (reviewed in ref. <sup>14</sup>), but low HIRs were found; soybean and tomato efforts have met with transformation and mutant complementation challenges. Here we report a crop *CENH3* HI line with commercial potential. The strategy to directly edit RFS alleles has several advantages over alternatives, including the use of a plant-preferred DNA repair pathway (non-homologous end joining) to avoid inefficient homology-directed repair.

The data are consistent with a hypothesis that *TaCENH3α* is a key factor in zygotic centromere formation<sup>18</sup>. Post-zygotic chromosome elimination is the likely mechanism of HI in our study, as in *Arabidopsis*<sup>5</sup>. Amino acids annotated as putative sites of histone modification were altered in RFS alleles, potentially affecting

nucleosome positioning<sup>19,20</sup> or kinetochore construction. Several results in our study contrast with established work in *Arabidopsis*. First, HI alleles are masked in heterozygous states of *Arabidopsis*, whereas wheat +/r induced seven-fold more haploids than r/r. However, a hypomorphic allele for the -A homoeolog is key: heterozygous (+/-)-A genotypes did not induce haploids. Second, the +/r, -/-, -/- G23 heteroallelic combination lowered seed set and exhibited strong segregation distortion against the RFS allele, suggesting that the A homoeolog is required during mega-gametogenesis. These results indicate that interaction between wild-type and RFS alleles in the inducer parent lowers female fecundity but also promotes a favorable reproductive state for HI. Although we did not determine whether a wild-type or r allele is present in embryo sacs that generate paternal haploids, expression of *CENH3α*-A during gametophytic development likely plays a key role. Assuming the r version of *CENH3* might lead to the formation of aberrant chromatin forms at centromeres, one would expect that centromeric chromatin in r embryos deriving from (+/r) parent plants might be less stable (due to +/r state exchange) than embryo sacs deriving from (r/r) parents, which have had a full generation to adapt to the r state. Further attention to expression and post-translational modifications of *CENH3α* and *CENH3β* in pre-meiotic germinal cells, meiocytes and female gametophyte development might clarify the specific genotypic states or other factors required to confer HI in each species.

Although maternal HI in wheat might be induced via wide cross with maize pollen<sup>21</sup> or knockout of *TaMATL*<sup>12,13</sup>, a paternal

HI system can be used for additional new breeding technologies, such as HI-Edit<sup>3</sup>, in which HI is used to edit elite crop genotypes (*CENH3* HI-Edit is four times more efficient than *MATL* HI-Edit). Another paternal HI technology is cytoplasm swapping<sup>22</sup>, which enables the discovery of mitochondrial or chloroplast genes that affect crop productivity<sup>23</sup>. For decades, *in vivo* nuclear transfer systems using paternal HI have been discussed in patent applications and literature reviews as a superior method of hybrid seed production. However, a technology to quickly convert female parent lines to male sterile cytoplasm has remained tantalizingly out of reach. Paternal HI would ensure inbred purity and reduce the cost of goods in corn, wheat and barley, which are reliant on traditional conversion involving 5–7 generations of backcrossing. Deployment of the *indeterminate gametophyte 1* (*ig1*) mutant<sup>24</sup> in corn has triggered lethal effects in some germplasms and an HIR of less than 1%. We also developed loss of function of *TaIG1* (TraesCS3A01G402300/TraesCS3B01G435700/TraesCS3D01G397200; *-/-*, *-/-*, *-/-*) but failed to obtain paternal haploids (0 of 228 F<sub>1</sub> plants; data not shown). The 7–8% HIR in this study mirrors the maternal HIR in *OsMATL* and *ZmMATL* knockouts; this is the first paternal HI system with commercial potential in crops. Considering conservation of *CENH3* in dicots and monocots, lessons drawn from this study increase the chances of success in other crops.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-020-0728-4>.

Received: 14 April 2020; Accepted: 5 October 2020;

Published online: 09 November 2020

### References

- Chang, M. T. & Coe, E. in *Molecular Genetics Approaches to Maize Improvement* (eds Kriz, A. L. & Larkins, B. A.) 127–142 (Springer, 2009)
- Jacquier, N. M. A. et al. Puzzling out plant reproduction by haploid induction for innovations in plant breeding. *Nat. Plants* **6**, 610–619 (2020).
- Kelliher, T. et al. One-step genome editing of elite crop germplasm during haploid induction. *Nat. Biotechnol.* **37**, 287–292 (2019).
- Ravi, M. & Chan, S. W. Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**, 615–618 (2010).
- Kuppu, S. et al. Point mutations in centromeric histone induce post-zygotic incompatibility and uniparental inheritance. *PLoS Genet.* **11**, e1005494 (2015).
- Kuppu, S., et al. A variety of changes, including CRISPR/Cas9-mediated deletions, in *CENH3* lead to haploid induction on outcrossing. *Plant Biotechnol. J.* **18**, 2068–2080 (2020).
- Kelliher, T. et al. Maternal haploids are preferentially induced by *CENH3*-tailswap transgenic complementation in maize. *Front. Plant Sci.* **7**, 414 (2016).
- Kelliher, T. et al. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **542**, 105–109 (2017).
- Gilles, L. M. et al. Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *EMBO J.* **36**, 707–717 (2017).
- Liu, C. et al. A 4-bp Insertion at *ZmPLA1* encoding a putative phospholipase a generates haploid induction in maize. *Mol. Plant* **10**, 520–522 (2017).
- Yao, L. et al. *OsMATL* mutation induces haploid seed formation in *Indica* rice. *Nat. Plants* **4**, 530–533 (2018).
- Liu, C. et al. Extension of the *in vivo* haploid induction system from diploid maize to hexaploid wheat. *Plant Biotechnol. J.* **18**, 316–318 (2020).
- Liu, H. et al. Efficient induction of haploid plants in wheat by editing of *TaMATL* using an optimized *Agrobacterium*-mediated CRISPR system. *J. Exp. Bot.* **71**, 1337–1349 (2020).
- Zhong, Y. et al. A DMP-triggered *in vivo* maternal haploid induction system in the dicotyledonous *Arabidopsis*. *Nat. Plants* **6**, 466–472 (2020).
- Kalinowska, K. et al. State-of-the-art and novel developments of *in vivo* haploid technologies. *Theor. Appl. Genet.* **132**, 593–605 (2019).
- Yuan, J., Guo, X., Hu, J., Lv, Z. & Han, F. Characterization of two *CENH3* genes and their roles in wheat evolution. *New Phytol.* **206**, 839–851 (2015).
- Karimi-Ashtiyani, R. et al. Point mutation impairs centromeric *CENH3* loading and induces haploid plants. *Proc. Natl. Acad. Sci. USA* **112**, 11211–11216 (2015).
- Maheshwari, S. et al. Naturally occurring differences in *CENH3* affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet.* **11**, e1004970 (2015).
- Jin, W. et al. Histone modifications associated with both A and B chromosomes of maize. *Chromosome Res.* **16**, 1203–1214 (2008).
- Deal, R. B. & Henikoff, S. Histone variants and modifications in plant gene regulation. *Curr. Opin. Plant Biol.* **14**, 116–122 (2011).
- Laurie, D. A. & Bennett, M. D. The production of haploid wheat plants from wheat × maize crosses. *Theor. Appl. Genet.* **76**, 393–397 (1988).
- Jaqueith, J. S. et al. Fertility restoration of maize CMS-C altered by a single amino acid substitution within the *Rf4* bHLH transcription factor. *Plant J.* **101**, 101–111 (2020).
- Flood, P. J. et al. Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *Nat. Plants* **6**, 13–21 (2020).
- Evans, M. The *indeterminate gametophyte 1* gene of maize encodes a LOB domain protein required for embryo sac and leaf development. *Plant Cell* **19**, 46–62 (2007).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

## Methods

**Vector construction.** *TaCENH3α* was amplified and sequenced from Fielder, a hexaploid soft white spring wheat line, for gRNA design. A pair of gRNAs was selected in the *CENH3α* target regions, and a cassette was synthesized by GenScript ([www.genscript.com](http://www.genscript.com)) including the gRNA scaffold and the TaU6 promoter. This cassette was cloned into a binary vector that included a PMI selectable marker and *Streptococcus pyogenes Cas9* gene driven using a sugar cane Ubiquitin promoter along with a double enhancer (eFMV-e35S). *SpCas9* was codon optimized for wheat, and nuclear localization signals were positioned at both the N and C terminal tails (Fig. 1b and Supplementary Fig. 1). gRNA1 locates in exon 1; gRNA2 locates in the intron 2/exon 3 splice site; gRNA3 locates in the junction region between exon 1 and intron 1; gRNA4 locates in the junction region between exon 4 and intron 4. gRNA1, gRNA2 and gRNA4 are conserved among 3A/B/D genes; gRNA3 is conserved only in A/D genes; there is one single-nucleotide polymorphism in the seed region of gRNA3 in the B gene.

**Wheat transformation and plant nursery.** Fielder wheat was used for transformation. Immature embryos were harvested at about 10 d after anthesis and sterilized with 70% ethanol for 1 min and 1% sodium hypochlorite for 10 min. Immature embryo transformation followed reported methods<sup>25</sup>: transformed plants were identified by selection medium containing mannose<sup>26</sup> and sampled for TaqMan analysis. Positive e35S and PMI events were sent to the greenhouse (Supplementary Table 7). T<sub>0</sub> plants were grown in 310 × 260-mm pots filled with peat, turf and perlite at a ratio of 4:4:1 plus 150 g of organic fertilizer per pot. Watering was managed via drip irrigation. Growth conditions were 22 ± 2 °C days and 18 ± 2 °C nights, and the photoperiod was set to 16-h days and 8-h nights.

**Tissue expression check of *TaCENH3α* and *TaCENH3β* homoeologs.** Total RNA from leaf, anther, pollen and ovaries of Fielder wheat was TRIzol extracted; 1 µg of total RNA was used for complementary DNA synthesis using the SuperScript III first-strand system (Invitrogen) and oligo-dT primers. Homoeolog-specific TaqMan assays were designed using Primer Express (sequences in Supplementary Table 7). The control assay for qRT was ADP-ribosylation factor. The relative expression of each homoeolog against the control was calculated via the 2<sup>-ΔΔCT</sup> method.

**Detection and interpretation of edits.** Plants carrying mutations in the A/B/D homoeologs of *CENH3α* were identified by sequencing. The targeted regions were amplified with KOD -PLUS- Neo (Toyobo) and cloned into the pEASY vector (pEASY-Blunt Zero Cloning Kit, TransGen Biotech). Ten independent random clones were selected for sequencing (Life Technologies). Sequences were aligned to the wild-type allele in Vector NTI.

**Screening of haploid F<sub>1</sub> plants using KASP assays.** Five KASP markers were used to distinguish 03S0352-22, a hard red spring wheat, from Fielder. The markers were in different chromosomes. KASP reagents were ordered from LGC.

**Ploidy check using flow cytometry.** One gram of fresh leaf tissue was sampled and chopped in 2 ml of LB01 extraction buffer<sup>27</sup>, filtered through an 80-µm membrane and centrifuged at 1,000 r.p.m. for 5 min. Isolated nuclei were then stained with propidium iodide for 20 min in the dark. A BD FACSCalibur flow cytometer was used to check the ploidy level. Data analysis was conducted using ModFit software. Wheat hybrid and inbred samples with known ploidy levels were used as a control. The results were displayed as a DNA histogram.

**Phenotyping.** The abaxial surface was torn using tweezers and put onto water on a clean glass slide. Stomata were checked and imaged using a light microscope. Plant height was measured from the soil surface to the tip of the highest spike.

Flowering stage was calculated from seed sowing to pollen shed of a primary spike. The number of tillers and length of the flag leaf were measured at the pollen shed stage. Pollen staining followed the described method<sup>1</sup>. Seed set data were captured from the outer two florets of each spikelet, excluding the two spikelets at the top and base of each spike.

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

## Data availability

The authors declare that all data supporting the findings of this study are available in the manuscript and the Supplementary Materials.

## References

25. Wright et al. Efficient biolistic transformation of maize (*Zea mays L.*) and wheat (*Triticum aestivum L.*) using the phosphomannose isomerase gene, *pmi*, as the selectable marker. *Plant Cell Rep.* **20**, 429–436 (2001).
26. Gui, H., Liu, Y., Han, K. & Li, X. The relationship between *PMI* (*manA*) gene expression and optimal selection pressure in *Indica* rice transformation. *Plant Cell Rep.* **33**, 1081–1090 (2014).
27. Loureiro, J., Rodriguez, E., Doležel, J. & Santos, C. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Ann. Bot.* **98**, 679–689 (2006).

## Acknowledgements

We thank Z. Chai, L. Geng, Y. Liu, X. Zhang, R. Cui and Y. Ren for their work generating the genome-edited plants. Thanks to Y. Ma and Y. Zhang for data analysis. We thank Y. Gao and W. Teng for wheat outcrossing and G. Tang for gene sequencing and characterization of *CENH3* expression. Thanks to M. Zong for ploidy analysis support and to W. Huang, C. Ma, Y. Pan, X. He, W. Jin, C. Gao and F. Han for technique suggestions. We thank J. Green, L. Kavanaugh, M. Rose and X. Tan for bioinformatics support and C. Leming for intellectual property guidance. We thank X. Zhang, D. Skibbe, X. Chen, J. Xu, K. White, T. Zhu, W. Cao, R. Quadt, B. Anindya, L. Shi, E. Dunder, I. Jepsen, G. Wu and B. Zhang for leadership and project guidance.

## Author contributions

J.L. and T.K. designed the experiments and wrote the manuscript. K.Y. managed plant care in the greenhouse, including seed increase and outcrossing. J.W. developed the genotyping strategy and conducted T<sub>0</sub>/F<sub>1</sub> genotyping and KASP check. C.L. did the vector design and construction. H.G. and Y.W. performed the wheat transformation. H.Z. and J.L. sequenced T<sub>0</sub>/T<sub>1</sub> genotypes. D.L. and J.L. performed pollen staining, stomata check and flower structure phenotyping, and R.C. performed the qRT-PCR of the three homoeologs of *TaCENH3α* and *TaCENH3β* in various tissues. T.L. analyzed histone modification sites. Q.Q., W.W., R.R. and X.Z. provided technique suggestions.

## Competing interests

A patent covering the information in this manuscript was submitted on 2 October 2018.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41587-020-0728-4>.

**Correspondence and requests for materials** should be addressed to J.L. or T.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Reporting Summary

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

### Statistics

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

n/a Confirmed

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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection The BD FACSCalibur™ flow cytometer was used to check ploidy level.

Data analysis Data analysis was conducted using the ModFit software.

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

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

There are no restrictions on data availability.

# Field-specific reporting

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

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	Sample sizes were chosen based on availability of seeds or plants, space considerations in the greenhouse and the desire to get statistically significant data to support meaningful conclusions.
Data exclusions	No exclusions were made of outliers.
Replication	Findings were replicated when necessary and possible; importantly, we tested the haploid induction rate via selfing and outcrossing (2 times) - and had replicates of 2 independent edited EO plants.
Randomization	Not relevant for this study because we didn't allocate samples into experimental groups, except when we sorted the haploids from the diploids. There is no area where randomization is relevant for this study.
Blinding	Not relevant - the only area this would be possible is analyzing physical traits of diploids and haploids - it was obvious which plants were haploids because they are so short, so it was impossible to blind the researcher.

## Reporting for specific materials, systems and methods

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

### Materials & experimental systems

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

### Methods

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

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

#### Sample preparation

One gram of fresh leaf tissue was sampled and chopped in 2ml of extraction buffer and filtered through an 80µm membrane and centrifuged at 1000 rev/min for 5min. Isolated nuclei were then stained with propidium iodide for 20min in the dark.

#### Instrument

The BD FACSCalibur™ flow cytometer was used to check ploidy level.

#### Software

Data analysis was conducted using the ModFit software.

## Cell population abundance

For this ploidy analysis experiment, cell sorting and purification were not involved. After filtration, cells in the suspension were performed experimental procedures. Around 5000 cells were analyzed per sample.

## Gating strategy

Overall, ploidy analysis is a rather simple use of flow cytometry, and so this section is not too complicated because the data collection and analysis is quite simple and straight forward compared to normal flow cytometry. You can see that our plots do not have multiple complex channel comparisons or statistics as they simply are not necessary. We are only trying to distinguish haploids from diploids and that is quite obvious to us and to reviewers / readers. The analysis was performed as described below:

Gating strategy is combining a scatter parameter (FSC/SSC) with a fluorescence parameter (FL2/SSC). The first step is to distinguish the cells based on their light scatter properties. We selected an area on the FSC/SSC scatter plots generated during the flow experiment and decided the cells we continue to analyze (Debris with small size could be distinguished by FSC which estimated the cell relative size, dead cell tend to have higher SSC). Then select an area on the FL2/SSC scatter plot. The different fluorescence signals of diploids and haploids allow them to be distinguished from cellular debris. Frankly, we didn't remove too much data points in gating, because the sample quality was good and the target population proportion was good enough to differentiate haploids and diploids. In order to identify the positive dataset (haploids), flow cytometry should be repeated in the presence of an negative control (diploids). The positive dataset was identified as a peak on the left with smaller FL2-A value, around 50. Negative control peak on the right with FL2-A around 100. We are not providing an example of the gating strategy because frankly, it was hardly used as the data was already quite clean and easy to interpret.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.