Article

Overriding Mendelian inheritance in *Arabidopsis* with a CRISPR toxin–antidote gene drive that impairs pollen germination

Yang Liu [1], Bingke Jiao [1,2], Jackson Champer [3] & Wenfeng Qian [1,2] ✉
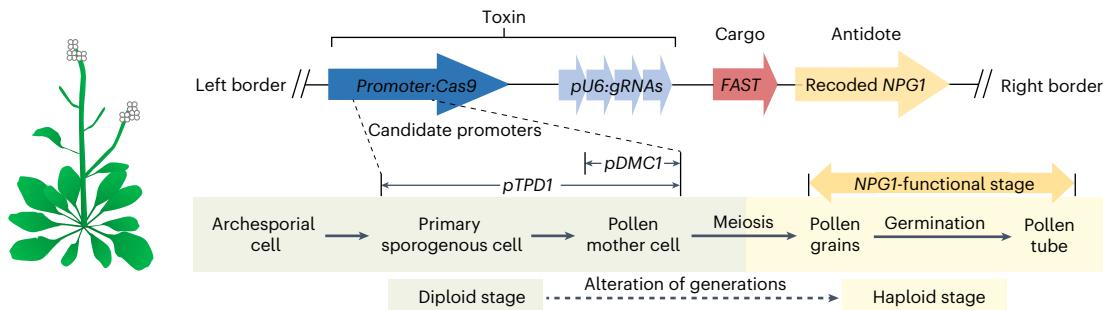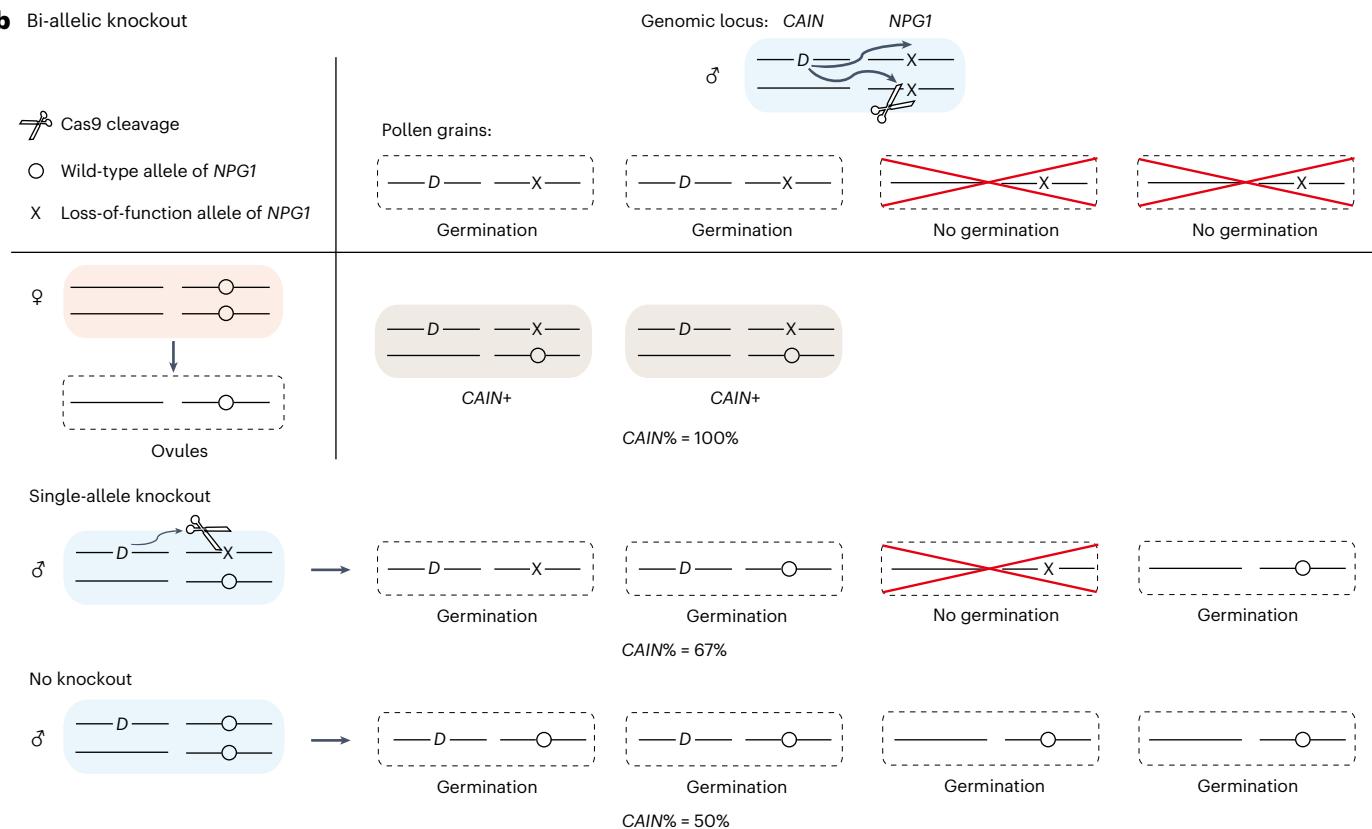
Synthetic gene drives, inspired by natural selfish genetic elements and transmitted to progeny at super-Mendelian (>50%) frequencies, present transformative potential for disseminating traits that benefit humans throughout wild populations, even facing potential fitness costs. Here we constructed a gene drive system in plants called CRISPR-Assisted Inheritance utilizing *NPG1* (*CAIN*), which uses a toxin–antidote mechanism in the male germline to override Mendelian inheritance. Specifically, a guide RNA–Cas9 cassette targets the essential *No Pollen Germination 1* (*NPG1*) gene, serving as the toxin to block pollen germination. A recoded, CRISPR-resistant copy of *NPG1* serves as the antidote, providing rescue only in pollen cells that carry the drive. To limit potential consequences of inadvertent release, we used self-pollinating *Arabidopsis thaliana* as a model. The drive demonstrated a robust 88–99% transmission rate over two successive generations, producing minimal resistance alleles that are unlikely to inhibit drive spread. Our study provides a strong basis for rapid genetic modification or suppression of outcrossing plant populations.

Facing diverse challenges such as threats to food security from agricultural weeds and the environmental crisis of invasive plants, the genetic manipulation of wild plant populations has emerged as a potentially powerful and transformative strategy. However, efforts to introduce traits that benefit humans into wild plant populations are almost invariably limited by classical Mendelian inheritance and Darwinian selection because such traits are probably detrimental to the plants themselves and therefore will be lost rapidly in targeted populations. Nevertheless, the presence of selfish genetic elements in plants, capable of transmitting to progeny at super-Mendelian (>50%) frequencies[1–8], provides a potential solution for manipulating wild plant populations. Inspired by these natural processes, synthetic gene drives have been proposed for spreading genetic alterations throughout populations in their natural environment, despite the potential fitness costs, providing a new direction to address agricultural and ecological challenges[9–12].

The most commonly implemented synthetic gene drive systems are homing-based drives, which have been shown to spread rapidly within populations in various non-plant hosts, including yeast[13], mosquitoes[14,15], flies[16] and mice[17]. This propagation is predominantly driven by the 'copy-and-paste' mechanism that depends on CRISPR–Cas9-mediated DNA cleavage (double-strand break (DSB)) of the homologous allele followed by homology-directed repair, thus converting diploid germline cells from heterozygotes into homozygotes. Nevertheless, homing-based drives frequently yield resistance alleles, defined as those that, because of the activity of endogenous end-joining DNA repair pathways, become resistant to further cleavage. Notably, this repair pathway is preferentially used in plants[18], thus posing a major challenge to sustained dissemination of homing-based gene drives in plant populations.

To develop an efficient synthetic gene drive for plants it is therefore necessary to circumvent dependency on the homology-directed

[1]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]Center for Bioinformatics, School of Life Sciences, Center for Life Sciences, Peking University, Beijing, China. ✉e-mail: wfqian@genetics.ac.cn

a Design of *CAIN*

b Bi-allelic knockout

Fig. 1 | The *CAIN* gene drive design and predicted genetic behaviour.
a, The T-DNA construct of *CAIN*. The corresponding stages of *Arabidopsis thaliana* male gametophyte development are depicted below. The *TPD1* promoter provides a longer window for the toxin to function in progenitor pollen cells, potentially facilitating higher overall Cas9 cleavage activity before *NPG1* function is first required in pollen germination. **b**, Anticipated fractions of F$_1$ progeny carrying *CAIN* (*CAIN%*) from a cross between a

CAIN-carrying male parent (*CAIN*/+) and a wild-type female parent (+/+), based on various scenarios of *NPG1* cleavage in the male germline. Salmon, sky blue and grey boxes represent female parents, male parents and progeny, respectively. Dashed boxes depict gametophytes. Red crosses represent pollen grains that cannot germinate. D, drive allele; X, loss-of-function allele of *NPG1*.

repair pathway. One solution is the utilization of toxin–antidote gene drives, mimicking naturally occurring toxin–antidote systems such as *S5* (ref. 4), *qHMS7* (ref. 7) and *RHS12* (ref. 8) loci in rice. These toxin–antidote systems typically feature a toxin, expressed before meiosis and consequently disseminated to all gametes, disrupting regular gametogenesis. The genetically linked antidote is activated in a later, postmeiotic stage to mitigate the toxin-induced damage, conferring a fitness advantage to its carrier. Although natural toxin–antidote systems involve sophisticated regulatory dynamics that might not seamlessly transfer across species, CRISPR–Cas9 could overcome this obstacle and provide a universal template for mimicking toxin–antidote strategies. In such a design, a CRISPR–Cas9 cassette could act as a toxin, cleaving an essential gene to produce a loss-of-function allele,

while a recoded, Cas9-resistant version of the essential gene functions as the antidote to rescue the defective functionality[19–21].

In this work, we present CRISPR-Assisted Inheritance utilizing *NPG1* (*CAIN*), a synthetic toxin–antidote gene drive developed for plants. We introduced *CAIN* into *Arabidopsis thaliana* and monitored its transmission through a red fluorescent seed phenotype, encoded within the drive. *CAIN* transmission rates greatly exceeded the expected Mendelian inheritance of 50% in heterozygous male parents, reaching 88–99% within two successive generations. Genotyping revealed a scarcity of resistance alleles, and modelling predicted that the drive could increase from 1% to 99% prevalence within roughly 17 outcrossed generations. The rapid spread of *CAIN* in *A. thaliana* demonstrates its potential for broad applications in diverse plant species, paving the

way for innovative approaches in ecological management and sustainable agriculture.

Results

Design of *CAIN*, a CRISPR-based toxin–antidote gene drive targeting pollen germination

CAIN consists of two tightly linked components: the toxin and the antidote, which can optionally be combined with a cargo. The toxin is a guide RNA (gRNA)–Cas9 cassette that can introduce loss-of-function mutations in a gene, *No Pollen Germination 1* (*NPG1*)[22], whose expression in pollen grains is essential for pollen germination. The antidote is a recoded, CRISPR-resistant version of *NPG1* (Fig. 1a and Extended Data Fig. 1). To mitigate potential negative effects on plant development and fitness caused by Cas9 activity in somatic tissues, we selected two promoters with distinct temporal expression patterns for male germline expression of *Cas9* (Fig. 1a): the *Disruption of Meiotic Control 1 (DMC1)* promoter, which expresses *Cas9* in the pollen mother cells within anthers[23], and the *Tapetum Determinant 1 (TPD1)* promoter, which activates gene expression in sporogenous cells that eventually develop into pollen mother cells[24]. To guarantee the effectiveness of *CAIN*'s toxicity, we incorporated four gRNAs (gRNA2, gRNA6, gRNA11 and gRNA23) of varying efficiencies into *CAIN* based on an initial assessment (Extended Data Fig. 1 and Supplementary Table 1), each expressed under the control of a U6 promoter[25]. As the cargo, we chose the red fluorescence protein marker, fluorescence-accumulating seed technology (FAST)[26], which generates a red fluorescent signal in dry seed, to facilitate phenotypic tracking of the gene drive (Fig. 1a). Furthermore, we generated a construct containing only the FAST marker to serve as a negative control (Extended Data Fig. 2).

In theory, the toxin can potentially disrupt both alleles of *NPG1* before meiosis and therefore impair germination of all four pollen grains, and the antidote can neutralize this effect only when present within a pollen grain. As a result, when a plant carrying a single copy of *CAIN* pollinates a wild-type plant, only the two *CAIN*-carrying pollen grains can successfully complete germination, yielding 100% transmission of *CAIN* (Fig. 1b). Even if only one of the two alleles of the essential gene is disrupted by the toxin, the transmission rate of *CAIN* will be two-thirds (Fig. 1b), well above 50% Mendelian inheritance.

This process can propagate in successive generations, ultimately disseminating *CAIN* throughout the population through continuous cross-pollination. Although gene drive systems are primarily designed to function in outcrossing populations, we selected *A. thaliana*, which primarily undergoes self-pollination, as our experimental model. Throughout the entire experiment, plants carrying *CAIN* were manually crossed with wild-type plants. Coupled with rigorous experimental procedures and management, the choice of this predominantly self-pollinating species further bolsters the biosafety of our gene drive.

Substantially increased transmission rate of *TPD-CAIN* to F_1 progeny

The *CAIN* design does not require linkage between the drive and the target gene, and consequently, we used *Agrobacterium*-mediated floral dipping to introduce the *CAIN* construct (*DMC-CAIN* and *TPD-CAIN*) into wild-type *A. thaliana* Columbia-0 (Col-0) at a random locus (Fig. 2a and Extended Data Fig. 3). We screened transformants (T_1) with the exogenous DNA sequence inserted into only a single locus (*DMC-CAIN*/+ or *TPD-CAIN*/+, where '+' represents the other allele that does not carry the drive) for subsequent characterization (Supplementary Table 2), because these transformants will exhibit a 50% transmission rate according to Mendelian inheritance in the absence of gene drive activity.

To evaluate whether *CAIN* can successfully enhance the likelihood of inheritance beyond Mendelian expectation through paternal transmission, we used *DMC-CAIN*/+ or *TPD-CAIN*/+ T_1 plants as the paternal donor in crosses with a wild-type female parent (Fig. 2a).

This experiment was effectively a test cross because the female parent would always supply a wild-type allele to the F_1 progeny, and therefore *CAIN* transmission from the male parent can be determined based on the dominant FAST (red fluorescence in F_1 seeds) phenotype. In the F_1 progeny of *DMC-CAIN*/+ T_1 plants, the fraction of individuals inheriting *DMC-CAIN* (*CAIN*%) was significantly higher than predicted Mendelian outcomes in only one (D31, *CAIN*% = 65.1%, false discovery rate = 0.004, binomial test) of the three crosses (Fig. 2b and Supplementary Table 3). By comparison, 89.6–96.9% of F_1 progeny in all four crosses of the *TPD-CAIN*/+ T_1 plants inherited *TPD-CAIN*, indicating a substantial deviation from Mendelian inheritance (Fig. 2b and Supplementary Table 3). By contrast, the negative control construct (FAST only) was indeed transmitted at a Mendelian rate of approximately 50% (Fig. 2b and Supplementary Table 3).

Because transmission rates in each silique (seed pod) can be treated as the result of an independent test, we further determined the significant deviation from a 50% transmission rate using replicated tests of goodness of fit (*G*-test, Fig. 2b) to account for variation in transmission rates among siliques. We found that the transmission rate of *DMC-CAIN* was significantly heterogeneous among siliques of male parent D31 ($P = 5 \times 10^{-5}$, heterogeneity *G*-test, Fig. 2b), which was the only cross showing a >50% transmission rate. This heterogeneity is probably caused by high stochasticity in Cas9 expression driven by the *DMC1* promoter, leading to non-trivial variability in *NPG1* cleavage status between pollen from different anthers or flowers. By contrast, no significant heterogeneity in transmission rate was detected among siliques of male parent *TPD-CAIN*/+ T_1 plants. Moreover, the F_1 populations of all four crosses showed *CAIN*% close to 100% (Fig. 2b), suggesting robust performance by *TPD-CAIN*.
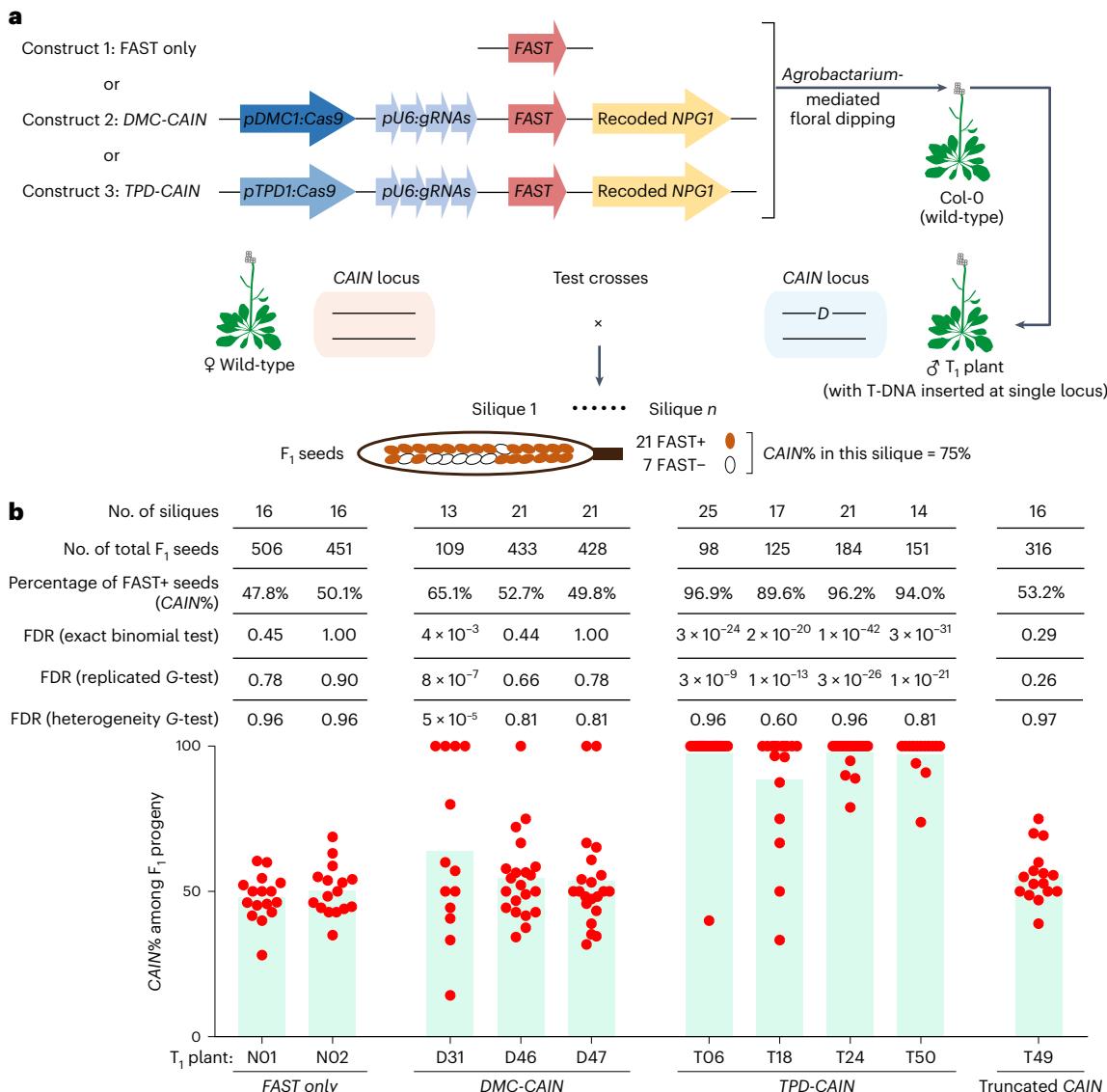
Transmission bias in *TPD-CAIN* is dependent on Cas9 activity

Previous research indicated that in the absence of Cas protein, gRNA could independently cause gene silencing in plants (gRNA-mediated gene silencing)[27]. This mechanism could theoretically contribute to biased inheritance, alongside Cas9-mediated *NPG1* disruption as designed (Fig. 1a). When performing transformation, we accidentally obtained a transgenic plant, T49, which lost the *TPD1* promoter crucial for Cas9 expression. When used as the male parent to pollinate wild-type plants, T49 showed Mendelian inheritance of the transgenic construct in its progeny (Fig. 2b), suggesting that the efficacy of the *CAIN* system is dependent on the activity of Cas9.

Transmission bias in *TPD-CAIN* is determined by direction of pollination

To assess whether *TPD-CAIN* biased Mendelian inheritance by disrupting *NPG1* as intended (Fig. 1), we genotyped F_1 plants by Sanger sequencing of the region around the Cas9 target sites in *NPG1* in somatic tissues (rosette leaves or early inflorescence) (Fig. 3a). Genotyping of FAST+ (*TPD-CAIN*/+) F_1 plants ($n = 49$) revealed that all carried a putative *NPG1* loss-of-function allele (*NPG1*⁻), with diverse indels observed at gRNA2 and gRNA11 target sites in 94% and 100% of F_1 progeny, respectively (Fig. 3b, Extended Data Fig. 4 and Supplementary Table 4). Theoretically, base substitutions and indels of length multiples of three could create CRISPR-resistant *NPG1* alleles yet preserve the reading frame, but they were rare (10% at the gRNA2 target site and 4% at the gRNA11 target site) (Supplementary Table 4). This rarity, especially across both gRNA sites (0%) (Supplementary Table 4), indicated that *TPD-CAIN* biased Mendelian inheritance probably through disrupting *NPG1* and consequently impairing pollen germination in pollen without *TPD-CAIN*.

Subsequent reciprocal crosses between *TPD-CAIN*/+ F_1 and the wild-type (+/+) were conducted to determine whether the biased inheritance of *TPD-CAIN* was influenced by the direction of pollination. When *TPD-CAIN*/+ F_1 plants served as the male parent ($n = 13$), *TPD-CAIN* transmission rates were significantly higher than expected in Mendelian inheritance (50%) (Fig. 4a and Supplementary Table 3). By contrast, in

Fig. 2 | Transmission rate of *CAIN* from the T$_1$ to F$_1$ generation. a, Experimental procedure for transformation yielding T$_1$ plants and subsequent test crosses generating F$_1$ progeny. The transmission rate of *CAIN* (*CAIN*%) was calculated from the fraction of F$_1$ seeds exhibiting FAST+. **b**, Bar plots showing the mean transmission rate estimated from individual siliques, for each pair of parent plants. Individual red dots indicate the transmission rate estimated from a single silique. Line T49 was incidentally transformed with truncated *TPD-CAIN* construct, with upstream -1.9 kb *TPD1* promoter lost and only the 5′-UTR remained. The null hypothesis for the two-sided exact binomial test and replicated tests of goodness of fit (*G*-test) was 50% transmission rate predicted by Mendelian inheritance. FDR, false discovery rate.
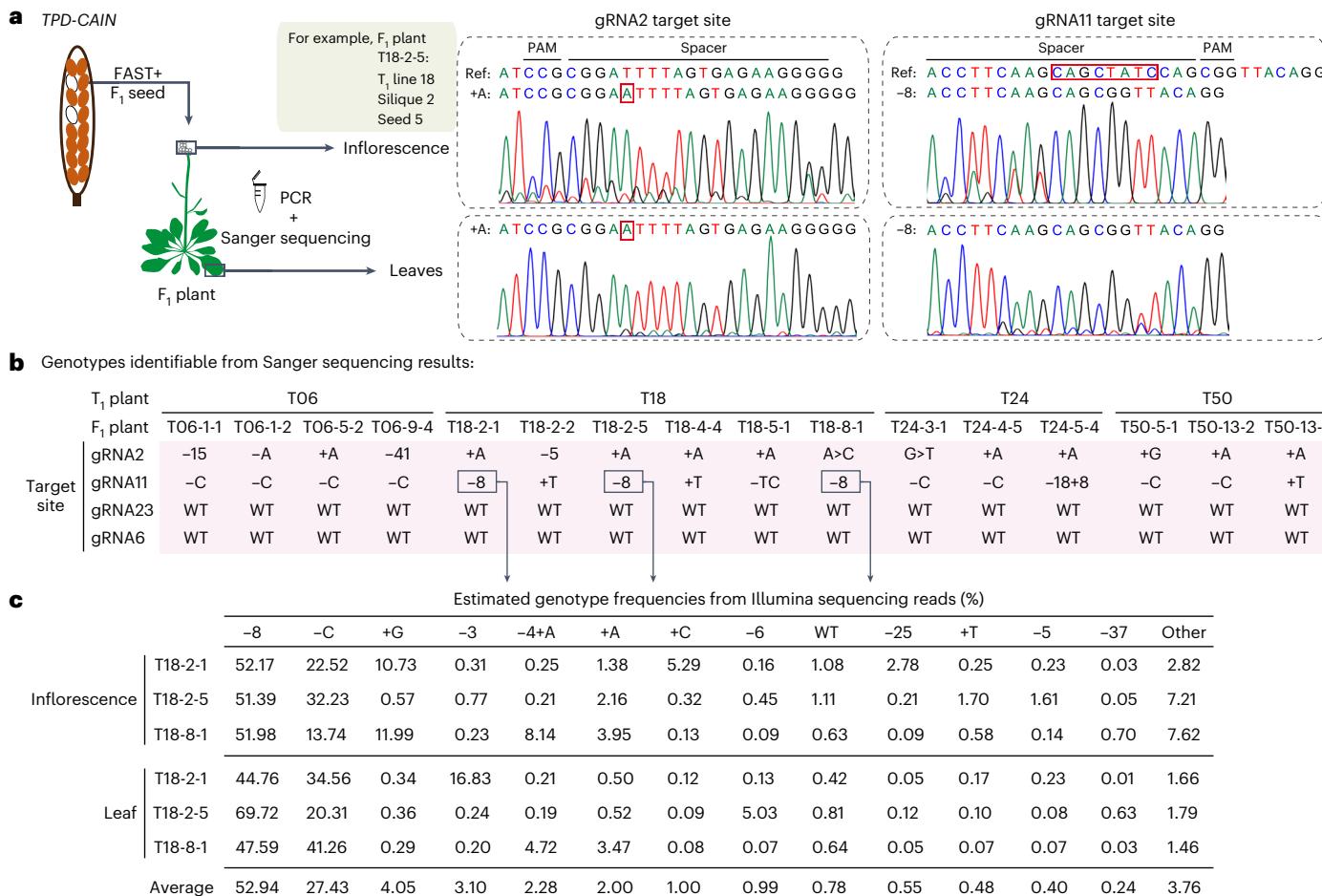
crosses with *TPD-CAIN*/+ F$_1$ plants serving as the female parent (*n* = 8), the *TPD-CAIN* transmission rate did not significantly deviate from 50% (Fig. 4b and Supplementary Table 3). These findings collectively indicated that the biased transmission of *TPD-CAIN* was probably because of a defective male gametophyte resulting from *NPG1* knockout.

The efficacy of *CAIN* is intrinsically tied to cleavage efficiency at the *NPG1* locus

Two potential explanations underlie the lower transmission rates of *DMC-CAIN* compared with *TPD-CAIN* (Fig. 2). First, Cas9 expression from the *DMC1* promoter may result in lower Cas9 activity than that driven by the *TPD1* promoter, and thus reduced DNA cleavage at the *NPG1* locus. Second, the timing of Cas9 activity might occur too late if expressed from the *DMC1* promoter, allowing sufficient expression of the native, undisrupted *NPG1* for the germination of some pollen grains. To differentiate between these hypotheses, we closely examined *NPG1* genotype in *DMC-CAIN*/+ F$_1$ plants (*n* = 18) (Supplementary Table 4).

Our analyses revealed that only two F$_1$ plants were *NPG1*$^{-/-}$ (D31-7-1 and D31-7-2), one plant was *NPG1*$^{+/-}$ (D31-8-3) and the remaining 15 plants were *NPG1*$^{+/+}$ (*NPG1*$^+$ representing the wild-type allele of *NPG1*) (Extended Data Fig. 5a,b and Supplementary Table 4). Note that the generation of *NPG1*$^{-/-}$ plants was probably a result of postfertilization Cas9 cleavage in F$_1$ plants (see below). These findings indicated that cleavage efficiencies were notably lower at the *NPG1* locus in *DMC-CAIN*/+ plants than in *TPD-CAIN*/+ plants (Fig. 3b), providing a plausible explanation for the varying transmission rates between the two *CAIN* constructs, without necessarily invoking timing-based explanation.

We next examined whether *DMC-CAIN*, despite its lower Cas9-mediated DNA cleavage efficiency, could still confer biased transmission across subsequent generations. For this purpose, we used two *DMC-CAIN*/+;*NPG1*$^{-/-}$ F$_1$ plants (D31-7-1 and D31-7-2) to pollinate wild-type plants, then screened for red fluorescence in F$_2$ seeds of both crosses (Extended Data Fig. 5b,c and Supplementary Table 3). *DMC-CAIN* transmission rates reached 95.9% and 99.5% in

Fig. 3 | Genotypes at the NPG1 locus for FAST+ F₁ plants generated from male T₁ plants carrying TPD-CAIN. a, Diagram of the somatic tissues sampled for genotyping F₁ plants, with example chromatographs. Red boxes denote insertions or deletions compared to the reference sequences. **b**, Sanger sequencing-based genotyping results for 16 FAST+ F₁ plants at four gRNA target sites. Complete genotyping data for all F₁ plants is available in Supplementary

Table 4. The symbols '+' and '−' before numerical numbers and bases denote insertion and deletion events, respectively. Numerical values indicate the number of nucleotides involved in the indels for cases in which the count exceeds two. The notation 'A>C' represents a base substitution from adenine (A) to cytosine (C). **c**, Illumina sequencing-based genotyping results for the gRNA11 target site. WT, wild-type allele.

the F₂ progeny of each respective cross. By contrast, the *CAIN*% was 63.7% (approximately two-thirds) in the F₂ progeny generated from the male parent *DMC-CAIN/+;NPG1⁺/⁻* F₁ plant D31-8-3 (Extended Data Fig. 5b,c and Supplementary Table 3). F₂ progeny of the remaining nine *DMC-CAIN/+;NPG1⁺/⁺* F₁ plants showed a *CAIN*% value close to 50% (Extended Data Fig. 5b,c and Supplementary Table 3). On average, the F₁ to F₂ transmission of *DMC-CAIN* is 57.5% (3,367 of 5,857) (Supplementary Table 3). This limited transmission rate underscored the crucial role of cleavage efficiency in determining the efficacy of the *CAIN* system.
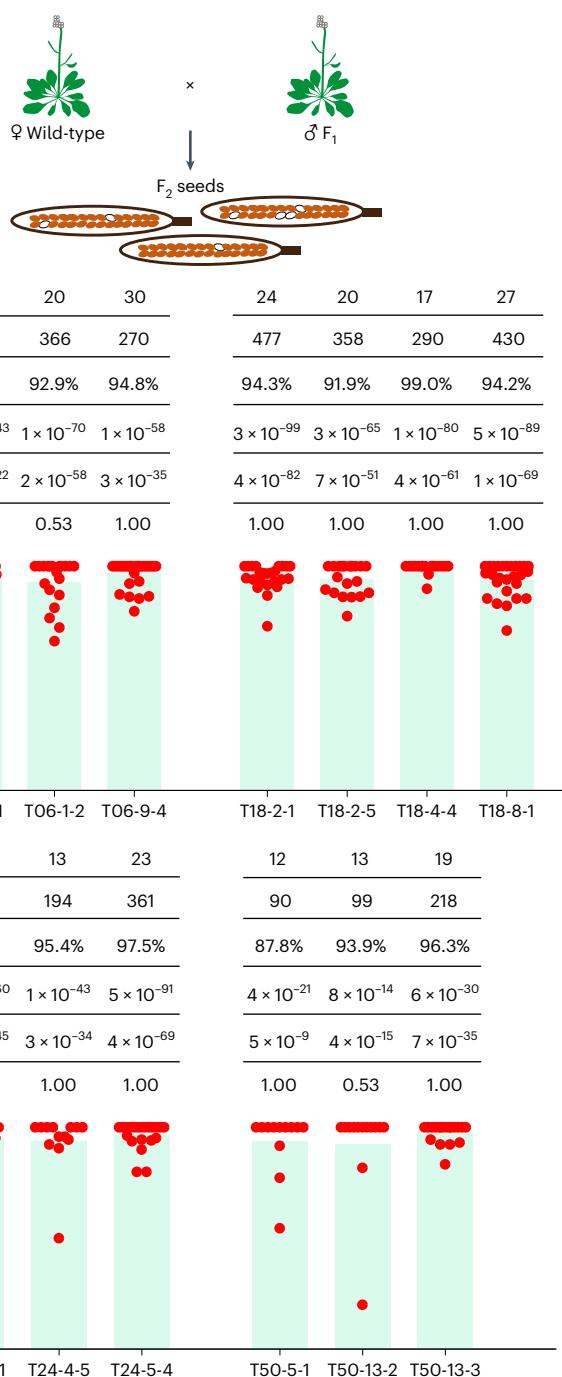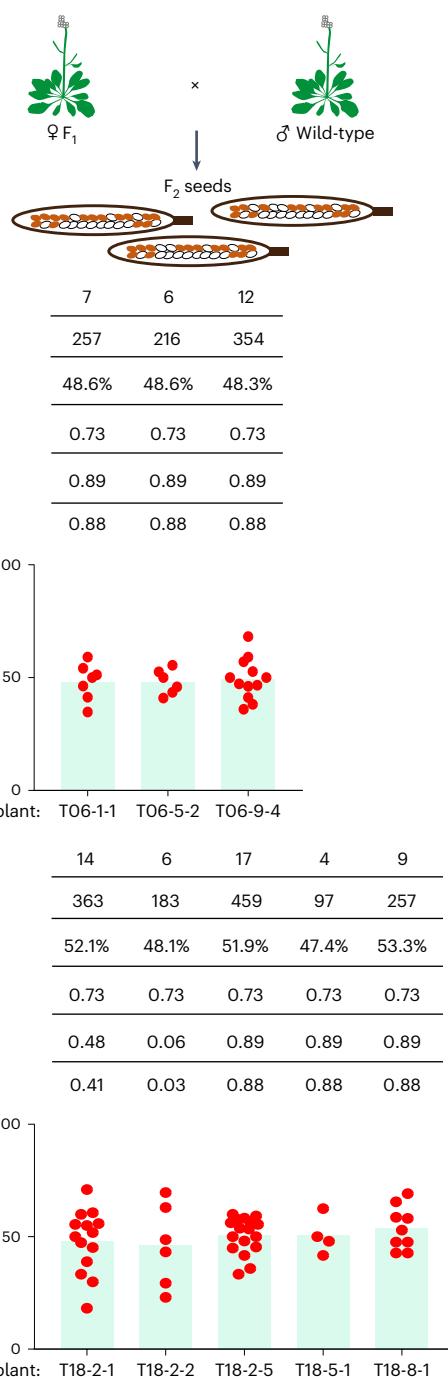
TPD-CAIN transmission rates <100% are attributable to failed DNA cleavage events and incomplete penetrance
Transmission rates of *TPD-CAIN* from male T₁ parents ranged from 89.6% to 96.9% (Fig. 2b and Supplementary Table 3), indicating that 3.1–10.4% of F₁ progeny did not inherit *TPD-CAIN* (FAST−), whereas 1.0–12.2% of F₂ progeny did not inherit the drive from male F₁ parents (Fig. 4a and Supplementary Table 3). This prompted our investigation into the causative factors resulting in a failure to inherit *TPD-CAIN*. Based on the relevance of *NPG1* cleavage in *DMC-CAIN* transmission, we hypothesized that FAST− F₁ plants might be the product of male gametophytes in which *NPG1* cleavage failed. We therefore scrutinized the genotypes of 11 FAST− F₁ plants at the *NPG1* locus in the region around the four gRNA target sites using Sanger sequencing (Fig. 5a). Notably, six of these plants were homozygous wild-type at all four sites (*NPG1⁺/⁺*)

(Fig. 5b), indicating that the male T₁ gametophyte contributed a *NPG1⁺* allele. These observations strongly suggested that multiple copies of *NPG1* evaded Cas9-mediated DNA cleavage in the male gametophyte of *TPD-CAIN/+* T₁ plants, despite successful integration of the drive.

By contrast, five of the FAST− F₁ plants had a *NPG1⁺/⁻* genotype at both the gRNA2 and gRNA11 sites (Fig. 5b), including three F₁ progeny of T18 and two F₁ progeny of T24. Considering that the female parent Col-0 can only contribute a *NPG1⁺* allele, the male parent should have provided the *NPG1⁻* allele through pollen transmission. It is noteworthy that postzygotic cleavage is an unlikely explanation for the presence of these *NPG1⁻* alleles because *TPD-CAIN* was absent in these FAST− F₁ plants. This finding suggested that some pollen grains carrying *NPG1⁻* alleles but lacking the antidote could still germinate, indicating incomplete penetrance of the *NPG1⁻* sterility phenotype. This could potentially result from a lack of Cas9 cleavage in some pollen mother cells, allowing sufficient *NPG1* expression for pollen grain germination before the wild-type *NPG1* was subsequently disrupted by Cas9 cleavage.

To understand the possible population-level impacts of failed DNA cleavage and incomplete penetrance on the dynamics of *CAIN* drive dissemination, we conducted computational simulations built on a Wright–Fisher model, assuming no fitness cost for *CAIN* carriers. This individual-based, stochastic model assumed a finite, randomly mating population reproducing in discrete, non-overlapping generations.
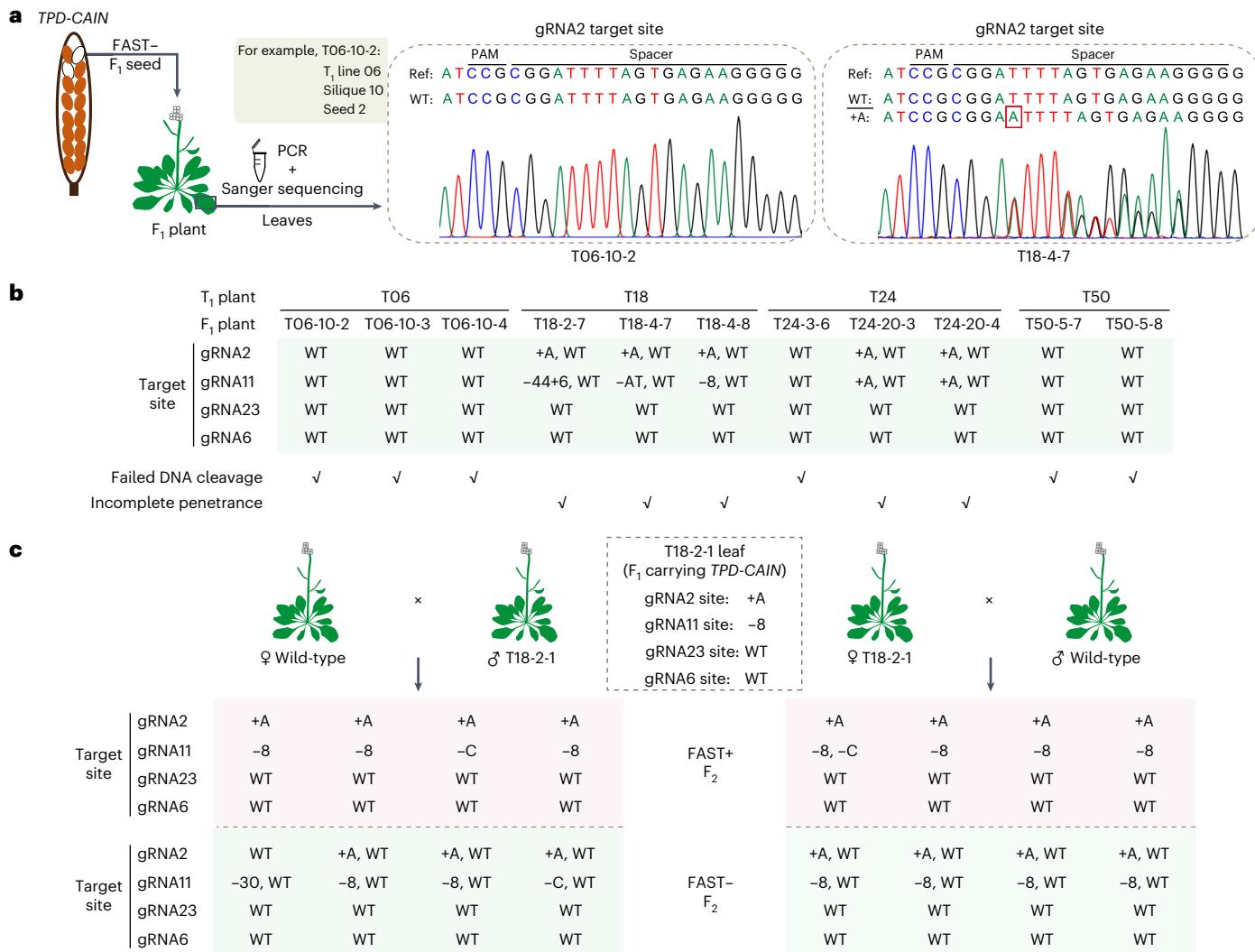
Fig. 4 | Transmission rate of *TPD-CAIN* from the F$_1$ to the F$_2$ generation in reciprocal crosses. a,b, Mean of *TPD-CAIN* transmission rate was depicted in bar plot with FAST+ F$_1$ plants serving as either the (**a**) male or (**b**) female parent.

We also assumed a high pollen count on stigmas and only one pollen donor per stigma, suggesting that a heterozygous *CAIN* carrier losing half its pollen would not impact the total offspring produced. Starting with 9,900 wild-type individuals and 100 *TPD-CAIN*/+ ones, we preserved population size, randomly chose mating pairs and confined DNA cleavage and transmission bias to male parents. Taking into account the factors of male germline cleavage efficiency and penetrance rate estimated for *TPD-CAIN* (98.4% and 96.0%, respectively) (Extended Data Fig. 6a), the simulation results indicated that its spread from 1% to 99% of the population would take approximately 17 outcrossed generations, only one generation longer than for optimal conditions (100% DNA cleavage efficiency in male germlines and complete penetrance)

(Fig. 6a), indicating that the efficiency of *TPD-CAIN* spread remained relatively robust despite these influences.

Prevalent somatic *NPG1* cleavage in *TPD-CAIN* carriers

A striking observation was the consistent absence of the *NPG1*[+] genotype detectable through Sanger sequencing in somatic tissues of FAST+ F$_1$ plants resulting from pollination of a wild-type plant (*NPG1*[+/+]) with *TPD-CAIN;NPG1*[−] pollens (Fig. 3b and Supplementary Table 4; *n* = 49). It suggested that the maternal *NPG1*[+] allele likely underwent CRISPR-mediated DSB. To further test this hypothesis, we conducted a thorough analysis of *NPG1* genotypes around the gRNA11 target site using Illumina sequencing on leaf and inflorescence samples from

Fig. 5 | Genotypes at the *NPG1* locus for FAST− F₁ and FAST+/− F₂ plants.
a, Diagram showing genotyping in the leaves of FAST− F₁ plants generated from male T₁ plants carrying *TPD-CAIN*. The red box denotes the inserted base compared to the reference sequence. **b**, Sanger sequencing-based genotyping results for 11 FAST− F₁ plants (generated from male T₁ plants carrying *TPD-CAIN*) at four gRNA target sites at the *NPG1* locus. The presence of check marks (√) for each F₁ plant indicates the inferred mechanisms—either failed DNA cleavage or incomplete penetrance. **c**, Sanger sequencing-based genotypes at four gRNA target sites for FAST+ and FAST− F₂ plants resulting from reciprocal crosses between F₁ plants carrying *TPD-CAIN* (*TPD-CAIN*/+) and the wild-type (+/+).
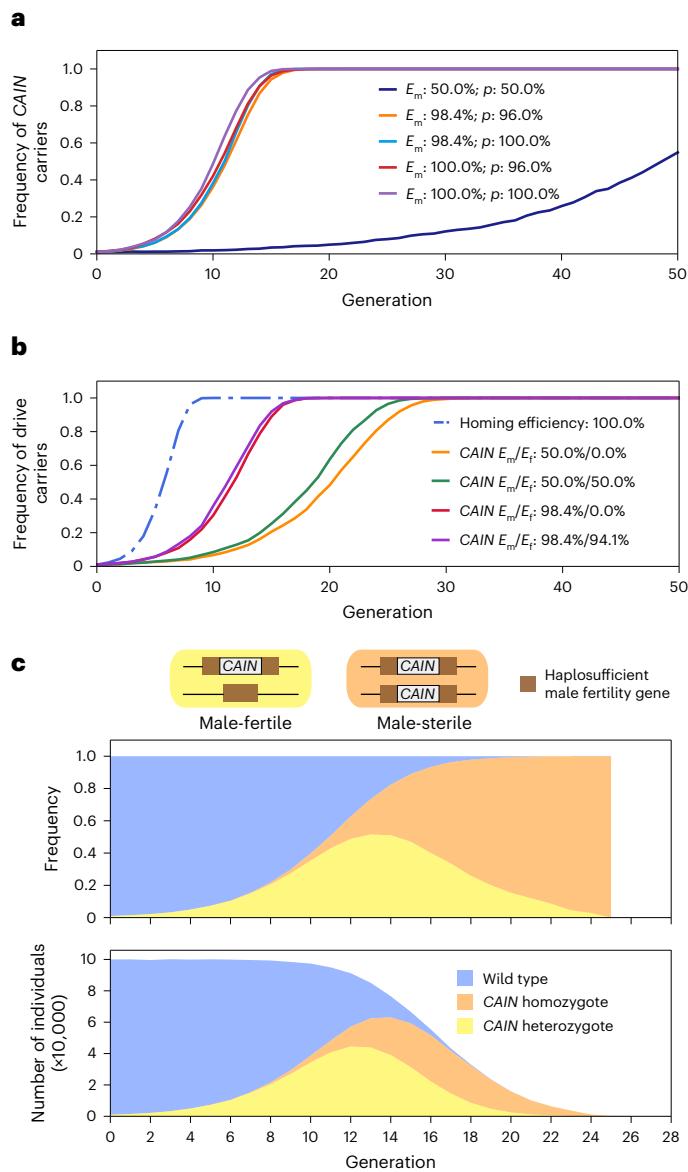
three FAST+ F₁ plants derived from male parent T18. Besides the primary (averaging 52.9%) genotype '*NPG1*⁻⁸', probably inherited from T18 (Fig. 3b), several other genotypes were also observed, although the un-cleaved *NPG1*⁺ allele was sparse (averaging 0.8%) (Fig. 3c). This diversity among *NPG1* alleles implied that DSBs in the maternal *NPG1* allele were induced by Cas9 after the one-cell zygote stage, with these DSBs primarily repaired via end-joining mechanisms in somatic cells.

The Cas9 activity responsible for inducing these DSBs could originate from one of the two plausible scenarios: *Cas9* expressed from the embryo's genome postfertilization, or the paternal carryover of Cas9 proteins. The latter scenario seemed less likely because of limited protein/RNA content allowed by the small sperm cell size[28]. To investigate these two hypotheses, we examined the *NPG1* genotype of F₂ plants generated by the reciprocal F₁ backcrosses conducted earlier (Fig. 4), using Sanger sequencing (Fig. 5c and Supplementary Table 5). The *NPG1*⁺ genotype seldom appeared in FAST+ F₂ progeny, regardless of *TPD-CAIN* inheritance from either the male or female parent, occurring only 2 of 17 times (gRNA2 target site) and 0 of 17 times (gRNA11 target site) when inherited from male parents, and 0 of 25 times at both gRNA2 and gRNA11 target sites when inherited from female parents (Supplementary Table 5). By contrast, the *NPG1*⁺ genotype was detected in all

FAST− F₂ progeny, even if the male parent carried *TPD-CAIN* (Fig. 5c and Supplementary Table 5). These observations aligned well with *NPG1* cleavage caused by postfertilization Cas9 expression, contradicting the explanation of paternal carryover. Note that these observations were also consistent with the known expression of *TPD1* in various plant tissues (leaves, flowers, ovule and carpel primordia)[24].

Enhancing *CAIN* dissemination through somatic *NPG1* cleavage

In plants, somatic mutations are often inherited because of the late differentiation of germlines[29]. As a consequence, somatic *NPG1* cleavage and subsequent repair has the potential to enhance the spread of *CAIN*, precluding the need for *NPG1* knockout later in the germline. To obtain an estimation of the overall apparent *NPG1* cleavage efficiency in female germlines, which we defined and applied consistently throughout this study as including cleavage in both germline and its precursor cells before differentiation, we analysed the genotypes of FAST− F₂ plants. These F₂ plants provided the genotypic information of the original *NPG1* alleles inherited from their parents, given that no additional *NPG1* cleavage would have occurred post-F₂ zygote formation. Among the 34 FAST− F₂ progeny resulting from the cross-pollination

Fig. 6 | Simulated spread dynamics of modification and suppression _CAIN_ drives. a, Computational simulation illustrating the impact of different settings for male germline cleavage efficiency (E_m) (empirical, 98.4%; artificial, 50.0% and 100.0%) and penetrance rate (_p_) (empirical, 96.0%; artificial, 50.0% and 100.0%) on the spread dynamics of _CAIN_. **b**, Computational simulation of the spread dynamics for the homing-based drive and _CAIN_, with an introduction frequency of 1%. The efficiency of the homing-based drive was set at 100% in both male and female germlines. For _CAIN_, the penetrance rate was set to the empirically estimated value of 96.0%. Various cleavage efficiencies were applied for male germline (empirical value of 98.4% or artificial value of 50.0%) and female germline (E_f) (empirical value of 94.1% and artificial values of 0.0% and 50.0%). **c**, The spread dynamics of the suppression _CAIN_ drive. The graph shows the changes in the number and frequency of _CAIN_ carriers and wild-type individuals across generations.

of FAST+ F$_1$ plants (_TPD-CAIN_/+) with wild-type pollen, 33 exhibited a heterozygous _NPG1_$^{+/−}$ genotype at the gRNA11 target site (Supplementary Table 5), higher than the 50% expected in the absence of somatic Cas9 cleavage. The prevalence of the _NPG1_$^{-}$ allele, probably introduced through the female gamete, suggested an apparent female germline cleavage efficiency of 94.1% (Extended Data Fig. 6b).

To assess how this female germline _NPG1_ cleavage could assist in _CAIN_ spread, we incorporated this parameter into a simulation with an initial introduction frequency of _CAIN_ carriers set at 1% (Fig. 6b). The results revealed a scenario-dependent impact: when male germline

cleavage efficiency is at 50%, an additional 50% female germline cleavage could expedite _CAIN_ propagation by approximately three generations. However, at the observed 98.4% male germline cleavage efficiency in _TPD-CAIN_ plants, the actual female germline cleavage efficiency of 94.1% only provided a marginal advancement of approximately one generation to _CAIN_ propagation (Fig. 6b). The simulation results also indicated that the rate of _CAIN_ spread was only delayed by a few generations compared with invasive, homing-based drives.

The fast propagation of _CAIN_ underscores its potential as a tool for population suppression. Unlike the approach used in population modification, where _CAIN_ is inserted at a random genomic location, population suppression can be attained by integrating _CAIN_ into a haplosufficient male fertility gene that is essential exclusively before the haploid stage. To gauge the effectiveness of this strategy, we conducted simulations on the population dynamics of _CAIN_ (Fig. 6c). These simulations were performed under the same assumptions as described earlier (Fig. 6a). The simulation results showed an initial rapid increase in both the number and frequency of _CAIN_ carriers, especially heterozygotes. Simultaneously, the total population size decreased progressively, ultimately resulting in population extinction by generation 26 (Fig. 6c), probably because of the surge in male-sterile _CAIN_ homozygotes. Collectively, these simulation results indicated that _CAIN_'s rapid propagation could allow it to be configured as a potent suppression drive.

Discussion

Mendelian inheritance, which is characterized by the equal probability of inheriting either allele from heterozygous diploid parents, promotes efficient adaptation through natural selection[30]. This same mechanism, however, prevents traits that are advantageous to humans yet harmful to the organisms themselves from spreading in wild populations. The _CAIN_ synthetic toxin–antidote gene drive system overrides Mendelian inheritance, thereby enabling the propagation of genes favoured by humans in wild plant populations.

Earlier synthetic toxin–antidote gene drives, such as synthetic Maternal Effect Dominant Embryonic Arrest (_Medea_)[31,32] and _Cleave and Rescue_ (_ClvR_)[19]/Toxin–Antidote Recessive Embryo (TARE)[20], were primarily designed to bias inheritance in female parents, and the resulting embryonic non-viability compromised fertility. The spread efficiency of these drives is limited by the proportion of individuals carrying the drive in the initial release into the wild population, necessitating higher release frequencies and rendering them inappropriate for population suppression[21]. By contrast, targeting male gametes, as in the proposed toxin–antidote dominant sperm (TADS)[21] and our _CAIN_ system, could mitigate potential fitness impacts on fecundity because of the higher abundance of haploid pollen grains compared with ovules or embryos (see detailed discussion below). Targeting male gametes potentially enhances the power of toxin–antidote systems, facilitating rapid drive propagation even at low frequencies.

Using the FAST marker[26] as the cargo gene, we established _CAIN_ as a state-of-the-art tool to efficiently modify entire plant populations. The cargo is versatile, adaptable for various ecological challenges, with gene selection dependent on specific scenarios and goals (Extended Data Fig. 7a). Alternatively, a _CAIN_ drive can operate by integrating into and consequently disrupting a specific gene. For instance, the rise of herbicide-resistant weeds often prompts increased herbicide usage[33,34], which in turn threatens biodiversity and potentially reduces productivity through broad spectrum effects. By inserting _CAIN_ into a herbicide resistance gene, the effectiveness of weed control can be maintained while reducing the need for excessive spraying, regardless of emerging resistance. This gene drive-based approach thus seeks to balance crop protection and environmental considerations to minimize the loss of biodiversity while optimizing productivity.

The implementation of _CAIN_/TADS gene drives in practical scenarios necessitates a thorough consideration of potential suppressor

mutations and their consequences. One major concern is the emergence of a cleavage-resistant, functional *NPG1* allele at the *NPG1* locus, which could prevent *CAIN* from spreading. To mitigate the risk of resistance alleles, we encourage future studies to use multiple, highly efficient gRNAs and target evolutionarily conserved genetic regions, where even mutations maintaining the reading frame could result in non-functional *NPG1* alleles.

Other types of suppressor mutations exist. For instance, the transposition or rearrangement of the recoded *NPG1* allele out of the *CAIN* locus could lead to an antidote-only suppressor. In population modification applications, loss-of-function mutations in the cargo, especially those with a fitness cost, could reverse population modification by eliminating cargo-carrying alleles. Note that loss-of-function mutations in Cas9 or the gRNA array are not considered major suppressor mutations in the *CAIN* system. This is because spread of the cargo can continue as long as there are other functional *CAIN* alleles in the population capable of disrupting *NPG1*. For population suppression, transposition of the wild-type male fertility gene to a different genomic locus could substantially hinder the population suppression efforts. Although these suppressors apparently reduce the power of the *CAIN*/TADS system, they would also make this gene drive self-limiting, generating only temporary or local impacts on the genetic composition of wild populations.

Note that in our simulations, we made several favourable assumptions (Fig. 6a–c). We assumed outcrossing and random mating, and no fitness cost for *CAIN* carriers. We also assumed an abundance of pollen grains on stigmas and a single pollen donor per stigma, implying that eliminating half the pollen from a heterozygous *CAIN* carrier would not affect the total number of offspring produced. However, factors such as inbreeding, fitness costs associated with *CAIN* and the fertility effects of pollen reduction can slow the spread of *CAIN* (Extended Data Fig. 8a–c). In addition, interactions between these factors, such as inbreeding with drive-related fitness costs, can further slow the spread of *CAIN* (Extended Data Fig. 8d). This highlights the need to consider specific life-history traits in wild plant species when implementing the *CAIN*/TADS system.

Acknowledging the ongoing intense debate and regulatory oversight of gene drive technologies[11,35–37], we prioritized biosafety in the development of *CAIN*. *A. thaliana* was selected as our self-pollinating model plant to prevent the inadvertent spread of *CAIN*, because gene drives require outcrossing to effectively spread throughout a population. The design of the *CAIN*/TADS system allows high specificity by enabling the targeting of distinct genotypes or ecotypes, achieved by selecting gRNA target sites based on naturally occurring genetic polymorphisms. The *CAIN*/TADS system is intrinsically zero-threshold, meaning it can spread throughout a population with just a few drive-carrying individuals, making it less confined compared with other CRISPR toxin–antidote drives like TARE[21]. Nevertheless, methods proposed to limit homing drive spread[38] can be integrated with *CAIN*. For example, by propagating a confined drive (for example, TARE or a variant thereof) to supply the required Cas9, the *CAIN* drive, with no *Cas9* included in its construct, can be 'tethered' to a specific geographic area by a TARE drive (Extended Data Fig. 7b). Such flexibility allows for a measured application of *CAIN*/TADS, offering tuneable levels of confinement for different scenarios and thus enabling secure application across diverse ecological contexts.

The scientific community, policymakers and stakeholders urgently need to engage in discussions about the responsible development and use of gene drive technologies. However, the implementation of self-containment strategies may not be feasible in cases of intentional misuse of gene drive technology, targeting domestic crops or wild plants. To mitigate this risk, it is crucial to develop safeguard technologies that can counteract the spread of gene drive alleles. One effective approach could be the intentional creation and if necessary, release of suppressor lines. Editing the native *NPG1* allele to resist Cas9 cleavage

is a particularly straightforward and efficient method. Another option is to engineer a new gene drive that can target the original cargo genes using the CRISPR–Cas9 system. Therefore, continuous monitoring of wild populations and the ability to rapidly produce suppressor strains are essential.

Our successful proof-of-concept demonstration of *CAIN* in *A. thaliana* sets a precedent for its application to other plant species, because *NPG1*, a critical component of our system, is highly conserved across a wide range of monocots and dicots[39]. For example, rice *OsPCBP*, homologous to *AtNPG1*, is transcribed in haploid pollen grains and is essential to mature pollen formation[40]. Looking forward, real-world application of the *CAIN* gene drive also faces other challenges, such as transgene silencing[41]. Future research in a particular plant species should consider optimizing factors like genomic integration sites, transgene characteristics and specific promoters to mitigate gene silencing and improve gene drive efficacy over generations. As we venture into this new frontier in genetic engineering, *CAIN*/TADS and other gene drive systems could reshape ecological management and agricultural practices.

Methods

Materials and growth conditions

All *A. thaliana* lines used in this study were derived from the Col-0 background. Seeds were surface sterilized with 10% sodium hypochlorite for 10 min, followed by three rinses with sterile water, before being sown on germination medium (2.2 g l^{-1} Murashige and Skoog medium, 10 g l^{-1} sucrose and 7.6 g l^{-1} plant agar, pH 5.7). After stratification at 4 °C for 2 days, the agar plates were transferred to a growth chamber set at 22 °C under a 16 h light/8 h dark photoperiod. Seven-day-old seedlings were then transplanted into soil and maintained in the greenhouse under the aforementioned conditions for further experiments.

Construction of plasmids

All restriction enzymes and enzymes for Gibson assembly were obtained from New England Biolabs. Hi-fidelity polymerase (Phanta Max Super-Fidelity DNA Polymerase, P505) for polymerase chain reactions (PCR) and the gel extraction kit (FastPure Gel DNA Extraction Mini Kit, DC301) were obtained from Vazyme. All the plasmids used in this study were cloned using standard molecular biology techniques and purified using StarPrep Fast Plasmid Mini Kit (Genstar, D201). All primers used in this study are listed in Supplementary Table 6.

Construction of the FAST only vector involved inserting the *FAST* marker[26] (*pOLE1:OLE1-TagRFP-Nos terminator*) into the XF675 binary vector through Gibson assembly[42]. To achieve this, the promoter and partial coding sequences of the *OLE1* gene (AT4G25140) were amplified from genomic DNA (gDNA), whereas *TagRFP* and *Nos terminator* sequences were synthesized as described in a previous study[26]. The HindIII-digested XF675 plasmid, along with the two fragments mentioned above, each having 20 bp overlaps with adjacent fragments, were assembled using Gibson assembly.

To generate *DMC-CAIN* and *TPD-CAIN* vectors containing *pDMC1/pTPD1:Cas9-4_gRNA_cassette-FAST-Rescue* (that is, *pNPG1:recoded NPG1*), all the components were cloned into the XF675 binary vector using Gibson assembly or golden gate ligation in five successive steps. First, the sgRNA cassette (*pU6-SmR-gRNA scaffold-U6 terminator*) was amplified from the plasmid pHEE401E (Addgene plasmid catalogue no. 71287) and inserted into the double-digested XF675 binary vector (digested with EcoRI and HindIII) using Gibson assembly. The HindIII recognition site was made complete for following digestion. Second, four gRNA cassettes, each driven by a U6 promoter (U6-26, U6-29, U6-1 and U6-26), were multiplexed using BsaI digestion and golden gate ligation as described[25]. Third, the resulting plasmid was then digested with HindIII, and the FAST marker sequences were cloned in using Gibson assembly, with the HindIII recognition site completed again. Fourth, the promoter (2,427 bp, amplified from gDNA) of *NPG1*, the recoded coding

sequence (without introns) obtained by PCR mutagenesis, the native 3′-untranslated region (UTR, 166 bp, amplified from complementary DNA) and the NOS terminator (amplified from the plasmid XF675) were subsequently inserted into the HindIII-digested plasmid using Gibson assembly. In this context, 'promoter' encompasses not just the classic promoter region preceding the transcription start site, but also includes the 5′-UTR and any introns located within the 5′-UTR. Fifth, the resulting plasmid was digested with KpnI, and the following fragments were cloned using Gibson assembly: the respective promoter sequences of *DMC1* (AT3G22880, 2,172 bp) and *TPD1* (AT4G24972, 2,770 bp) amplified from gDNA, the plant-codon-optimized *SpCas9* sequence amplified from the plasmid pHEE401E and the NOS terminator sequences amplified from the plasmid XF675.

Design principles for *CAIN*

Implementing a successful toxin in *CAIN* requires a gene to be essential for pollen germination. To identify suitable target genes, we retrieved candidates from a previously compiled list of genes important for male gametophyte development, but which do not affect female gametophyte development[43]. *No Pollen Germination 1* (*NPG1* or AT2G43040), which is essential for pollen germination[22], was selected as the target for our gRNA–Cas9 cassettes because its later role in male gametophyte functionality should allow more time for CRISPR-based gene knockout to occur compared with genes with earlier stage functions (Fig. 1a). The antidote was designed using a recoded version of *NPG1*, driven by its native promoter, to restore normal pollen germination. To ensure resistance to further Cas9 cleavage, we introduced 28 synonymous mutations around the four gRNA target sites (Extended Data Fig. 1a). In addition, introns were removed from the recoded version to ensure that *CAIN* constructs were as small as possible (Extended Data Fig. 1b,c).

The design of the *CAIN* system does not require a genetic link with its target gene, *NPG1*. In scenarios of low Cas9 cleavage efficiency, such a genetic link could actually decelerate the *CAIN* system's spread (Extended Data Fig. 8e). This is probably because the linkage hinders the enrichment of *CAIN*;*NPG1*⁻ plants in the subsequent generation.

Evaluating the editing efficiency of potential target sites in protoplasts

The CRISPR-P 2.0 (ref. 44) bioinformatics tool (http://crispr.hzau.edu.cn/CRISPR2/) identified 35 potential Cas9-mediated DNA cleavage target sites within the *NPG1* coding sequence. These were further refined based on three criteria: (1) positioning within the first two-thirds of the coding sequence; (2) having fewer than four consecutive thymine bases within the gRNA sequence; and (3) to avoid unintended secondary structures in guide RNA (including both the 20-nucleotide single guide RNA and the longer scaffold RNA), the sequence of the single guide RNA should not have more than seven-base pairing with the subsequent scaffold RNA sequence, predicted by CRISPR-GE[45] (http://skl.scau.edu.cn/home/). To evaluate the editing efficiencies of the remaining 12 gRNA target sites in *A. thaliana* protoplasts (Supplementary Table 1), the 20-nucleotide gRNA spacer sequences were individually incorporated into the *pAtU6-sgRNA* vector (Addgene plasmid catalogue no. 119775) using BsaI digestion and golden gate ligation.

A. thaliana protoplasts were prepared following procedures described in previous studies[46,47]. After co-transformation with *pAtU6-sgRNA* and *p2X35S-Cas9* (Addgene plasmid catalogue no. 53064), protoplasts were cultivated at room temperature for 48 h before harvest. In parallel, another set of protoplasts was transformed with *p2X35S-GFP* alone to measure the transformation efficiency approximately 16 h post transformation. The transformation efficiency was found to be 41% and 45% in two biological replicates, respectively.

gDNA was extracted from each tube of protoplasts using Plant Genomic DNA Kit (DP305, Tiangen). The genomic regions spanning 180–220 bp around each target site were PCR amplified and purified. To differentiate PCR products from two biological replicates at the same

target site, a six-nucleotide barcode (ATGCAG) was introduced for one of the replicates using primers (Supplementary Table 6). Upon purification, each of the 24 PCR product samples was quantified using NanoDrop 2000, pooled in equal amounts, and then subjected to library construction and Illumina paired-end 150 bp (PE150) sequencing (NovaSeq 6000).

A total of 7,134,194 sequencing reads were obtained from the 24 PCR products. After filtered using fastp[48] (v.0.23.1) with the parameter 'fastp -g -q 5 -u 50 -n 15 -l 150 --min_trim_length 10 --overlap_diff_limit 1 --overlap_diff_percent_limit 10', 7,055,416 clean reads were retained. The data for the two biological replicates were partitioned according to the presence or absence of the barcode sequence 'ATGCAG', and then mapped to *NPG1* reference sequence using Burrow–Wheeler Aligner[49] (BWA, v.0.7.17-r1188) with the parameter 'bwa mem -M'. Variants were called from the BAM files using SAMtools[50] (v.1.13) with the parameters 'samtools mpileup -A -d 0 -Q 13 -q 30 -B'. The depth of reads covering the 12 gRNA regions spanned from 57,412 to 163,671 in replicate 1 and from 133,643 to 203,234 in replicate 2, respectively. The mpileup files were then transformed into VCF files using an in-house perl script (https://github.com/QianLabWebsite/GeneDrive).

To identify CRISPR-mediated editing events, we considered variants exhibiting mismatches in the 23-nucleotide spacer and PAM region. Although the mismatches could potentially arise from Illumina sequencing errors, most identified mismatches were indels. By contrast, Illumina sequencing errors usually result in single base substitutions (10^{-3} mismatches per base[51]) rather than indels (10^{-6} mismatches per base[52]). To further mitigate errors in variant identification, a frequency threshold of 0.5% was set of single base substitutions and all indels were identified as edited variants. The editing efficiency for each gRNA was estimated by calculating the ratio of mismatched reads to the total number of reads mapped to the respective target site (Supplementary Table 1).

Generation of single-locus transfer DNA inserted transformants

Wild-type Col-0 plants were subjected to the floral dipping transformation[53] using *Agrobacterium* GV3101 strain containing one of the three vectors (*DMC-CAIN*, *TPD-CAIN* or FAST only). Successful primary transformants (T_1) were selected based on the presence of red fluorescence in dry seeds, detected by a handheld fluorescence detector (LUYOR 3415RG).

To identify single-locus T-DNA inserted transformants, thermal asymmetric interlaced-PCR[54] was conducted on 48–50 randomly chosen T_1 plants from each drive construct. Further validation was carried out using whole-genome sequencing (~20× coverage, NovaSeq 6000 PE150, Novogene) on candidate T_1 lines. We developed an in-house pipeline for detecting T-DNA insertion sites. We first created a reference file by concatenating T-DNA sequence with the *A. thaliana* reference genome[55] (TAIR 10). Sequencing reads were then aligned to this reference using BWA with parameters 'bwa mem -M'. We examined chimeric reads (reads with subsections aligning to separate positions in the reference genome) and discordant read-pairs (pairs with unexpected distance/orientation in their alignments) mapped to the T-DNA borders (300 bp). The residual sequences in these reads (or read-pairs) helped us pinpoint putative insertion sites in the genome. We performed local assembly of reads around each putative insertion site using Megahit[56] (v.1.2.9) with parameters 'megahit --prune-level 1 --prune-depth 1 --low-local-ratio 0.1', and the resulting contigs were aligned to the integrated reference with Blastn[57] (v.2.9.0+). This allowed us to reconstruct the junction between T-DNA and the chromosome at the insertion site. Finally, we confirmed single-locus T-DNA insertion in three *DMC-CAIN*, four *TPD-CAIN* and two FAST only T_1 lines (Supplementary Table 2).

Cross-pollination and the estimation of transmission rate

We assessed the transmission rate of the *CAIN* gene drives and the FAST only construct through cross-pollination experiments, quantifying

the percentage of progeny seeds that displayed red fluorescence. For cross-pollination, we transferred pollen from two or three male parent flowers to the stigma of an unopened, emasculated female parent flower.

R (v.4.2.3) and RStudio software (v.2023.03.0+386) were used for statistical analyses. We assessed deviations from Mendelian inheritance by examining the fraction of FAST+ seeds through a two-sided exact binomial test, setting a null hypothesis of 50%. Considering potential heterogeneity in *NPG1* cleavage status across male parent flowers, and consequent variations in the proportion of FAST+ seeds across individual siliques, we employed a replicated *G*-test from the DescTools package (v.0.99.49) to detect distortion in the transmission ratio. To account for an increased risk of type I error (false-positive) because of multiple comparisons, we calculated the false discovery rate using the p.adjust() function.

Genotyping

The genotypes at the *NPG1* locus were determined by Sanger sequencing of a ~2 kb PCR product, amplified from gDNA and covering all four gRNA target sites (Extended Data Fig. 1c; primers detailed in Supplementary Table 6). DNA was extracted from rosette leaves and early inflorescence (with flowers unopened). Consistent Sanger sequencing results were usually obtained from leaf and inflorescence samples of the same plant. However, if multiple peaks were evident in the chromatograph of leaf samples (indicating potential mosaicism or heterogeneity), genotypes were determined using inflorescence sample from the same plant. Genotyping of FAST− F_1 and F_2 plants was performed using only leaf samples; samples with indeterminable genotypes were excluded.

For three F_1 plants, we determined the gRNA11 target site genotypes using Illumina sequencing on both leaf and inflorescence samples. PCR products from different tissues, each with a unique barcode introduced during PCR (Supplementary Table 6), were mixed in equal quantities and sequenced with Illumina PE150. Clean reads, sorted by barcode, were mapped to the genomic sequence around the gRNA11 target site using BWA. The depth of reads covering the gRNA11 region varied between 107,207 and 150,422. We considered variants with mismatches in the 23-nucleotide spacer and PAM region, and determined the editing efficiency per sample as the ratio of reads with mismatches to total reads mapped to the target site. We identified single base substitutions with a frequency above 0.5% and all indels as edited variants.

CRISPR cleavage potential beyond *NPG1*, caused by the incorporated gRNAs in *CAIN*, raises concerns about off-target effects, which could unintentionally introduce genomic mutations and increase the genetic load to the population. We predicted 16 potential off-target sites using CRISPR-P 2.0 (ref. [44]) and tested the genotypes at these locations in 16 FAST+ F_1 plants (*TPD-CAIN*/+;*NPG1*$^{-/-}$) using Sanger sequencing (Supplementary Table 7). The absence of mutations at these potential off-target sites in all the tested plants indicated that the four gRNAs in *CAIN* specifically targeted *NPG1* as designed.

Simulation of the population dynamics

The population dynamics of *CAIN* were computationally simulated using an individual-based stochastic model based on the Wright–Fisher model, which assumes a constant population size and non-overlapping generations. For the modification *CAIN* drive, we considered genotypes at two unlinked loci, *CAIN* and *NPG1*, and initiated the population with 9,900 wild-type individuals and 100 *CAIN*-carrying heterozygotes (*TPD-CAIN*/+;*NPG1*$^{+/+}$). We implemented a density regulation production strategy following a previous study[21]. Briefly, we calculated a scaling factor (*S*) using the formula $S = 10/(9 \times N/K + 1)$, where *N* is the population size of the current generation, and *K* is the environmental carrying capacity (10,000). We calculated the offspring number (*X*) for each cross from a binomial distribution $X \sim B(50, 0.04 \times S)$. For each cross, a pair of individuals was randomly

chosen to produce offspring, with the roles of male and female randomly assigned. This process was repeated *N*/2 times, and the resulting offspring replaced the parent generation.

For each generation, male parents carrying *CAIN* could trigger CRISPR-mediated cleavage at the *NPG1* locus, at a rate either estimated from empirical data (98.4%) (Extended Data Fig. 6a) or set artificially (50% or 100%). This cleavage led to the creation of loss-of-function *NPG1* alleles, and the corresponding lethal phenotype in male gametes had a penetrance rate either empirically determined (96.0%) (Extended Data Fig. 6a) or artificially set (100%). We also considered cleavage efficiency in *TPD-CAIN*/+ females, set either based on empirical data (94.1%) (Extended Data Fig. 6b) or artificially (0%, 50%, 100%). For the homing-based drive, the presence of the drive allele in a heterozygous state led to the conversion of the wild-type allele into the drive allele with a 100% efficiency in both male and female germlines.

For the suppression *CAIN* drive, *CAIN* was placed within and consequently disrupted a haplosufficient male fertility gene[21]. Thus, the male homozygous *CAIN* plants were incapable of producing viable pollen grains. Note that when the *CAIN* construct is inserted into a female fertility gene or viability gene instead of a male fertility gene, the population would reach equilibrium without complete collapse, as illustrated in Extended Data Figure 8f,g, probably because of a failure to remove heterozygous individuals from the population. These simulations were performed under the same assumptions as that of the modification *CAIN* drive, with the exception that the initial population size and *K* were both set to 100,000.

To simulate the impact of inbreeding on the dynamics of *CAIN* spread, we established varying levels of inbreeding, denoted as (*f*). In this model, selfing—in which both male and female gametes originate from the same individual—happens at a frequency determined by *f*, whereas outcrossing takes place at a complementary frequency of $1 - f$. To account for the fitness cost for *CAIN* carriers, we applied relative fitness values as weight coefficients in the selection of parents for mating, using an additive genetic model (that is, without dominance). Furthermore, to represent the impact of diminished pollen production in male heterozygous *CAIN* carriers on reproduction, we introduced a 'fertility value'. This value ranges from 0.5, indicating a single offspring resulting from a mating event, to 1.0, indicating two offspring. For example, a fertility value of 0.7 means there is a 60% chance that a mating event will produce one offspring and a 40% chance it will result in two offspring.

The simulation was implemented using a custom Python script, available at https://github.com/QianLabWebsite/GeneDrive.

Reporting summary

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

Data availability

The Illumina sequencing data have been deposited in the Genome Sequence Archive[58] in National Genomics Data Center[59], China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences under accession number CRA011573. Plasmid sequences have been deposited in GenBase (https://ngdc.cncb.ac.cn/genbase/?lang=en, accession ID: C_AA031173.1, C_AA031174.1 and C_AA031175.1).

Code availability

The scripts developed for variant calling at gRNA target sites and simulation of the population dynamics for *CAIN* are available at GitHub (https://github.com/QianLabWebsite/GeneDrive).

References

1. Rhoades, M. M. Preferential segregation in maize. *Genetics* **27**, 395–407 (1942).

2. Fishman, L. & Saunders, A. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science* **322**, 1559–1562 (2008).

3. Long, Y. et al. Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proc. Natl Acad. Sci. USA* **105**, 18871–18876 (2008).

4. Yang, J. et al. A killer-protector system regulates both hybrid sterility and segregation distortion in rice. *Science* **337**, 1336–1340 (2012).

5. Shen, R. et al. Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. *Nat. Commun.* **8**, 1310 (2017).

6. Koide, Y. et al. Lineage-specific gene acquisition or loss is involved in interspecific hybrid sterility in rice. *Proc. Natl Acad. Sci. USA* **115**, E1955–E1962 (2018).

7. Yu, X. et al. A selfish genetic element confers non-Mendelian inheritance in rice. *Science* **360**, 1130–1132 (2018).

8. Wang, C. et al. A natural gene drive system confers reproductive isolation in rice. *Cell* **186**, 3577–3592 (2023).

9. Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. B* **270**, 921–928 (2003).

10. Sinkins, S. P. & Gould, F. Gene drive systems for insect disease vectors. *Nat. Rev. Genet.* **7**, 427–435 (2006).

11. Champer, J., Buchman, A. & Akbari, O. S. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* **17**, 146–159 (2016).

12. Bier, E. Gene drives gaining speed. *Nat. Rev. Genet.* **23**, 5–22 (2022).

13. DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. & Church, G. M. Safeguarding CRISPR–Cas9 gene drives in yeast. *Nat. Biotechnol.* **33**, 1250–1255 (2015).

14. Windbichler, N. et al. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**, 212–215 (2011).

15. Li, M. et al. Development of a confinable gene drive system in the human disease vector *Aedes aegypti*. *eLife* **9**, e51701 (2020).

16. Gantz, V. M. & Bier, E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **348**, 442–444 (2015).

17. Grunwald, H. A. et al. Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* **566**, 105–109 (2019).

18. Mao, Y. et al. Application of the CRISPR–Cas system for efficient genome engineering in plants. *Mol. Plant* **6**, 2008–2011 (2013).

19. Oberhofer, G., Ivy, T. & Hay, B. A. Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proc. Natl Acad. Sci. USA* **116**, 6250–6259 (2019).

20. Champer, J. et al. A toxin–antidote CRISPR gene drive system for regional population modification. *Nat. Commun.* **11**, 1082 (2020).

21. Champer, J., Kim, I. K., Champer, S. E., Clark, A. G. & Messer, P. W. Performance analysis of novel toxin–antidote CRISPR gene drive systems. *BMC Biol.* **18**, 27 (2020).

22. Golovkin, M. & Reddy, A. S. A calmodulin-binding protein from *Arabidopsis* has an essential role in pollen germination. *Proc. Natl Acad. Sci. USA* **100**, 10558–10563 (2003).

23. Klimyuk, V. I. & Jones, J. D. *AtDMC1*, the *Arabidopsis* homologue of the yeast *DMC1* gene: characterization, transposon-induced allelic variation and meiosis-associated expression. *Plant J.* **11**, 1–14 (1997).

24. Yang, S. L. et al. *Tapetum determinant1* is required for cell specialization in the *Arabidopsis* anther. *Plant Cell* **15**, 2792–2804 (2003).

25. Xing, H. L. et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 327 (2014).

26. Shimada, T. L., Shimada, T. & Hara-Nishimura, I. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. *Plant J.* **61**, 519–528 (2010).

27. Sharma, V. K. et al. CRISPR guides induce gene silencing in plants in the absence of Cas. *Genome Biol.* **23**, 6 (2022).

28. Zou, J. et al. Comparative proteomic analysis of *Arabidopsis* mature pollen and germinated pollen. *J. Integr. Plant Biol.* **51**, 438–455 (2009).

29. Lanfear, R. Do plants have a segregated germline? *PLoS Biol.* **16**, e2005439 (2018).

30. Crow, J. F. Genes that violate Mendel's rules. *Sci. Am.* **240**, 134–147 (1979).

31. Chen, C. H. et al. A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **316**, 597–600 (2007).

32. Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T. & Akbari, O. S. Synthetically engineered *Medea* gene drive system in the world-wide crop pest *Drosophila suzukii*. *Proc. Natl Acad. Sci. USA* **115**, 4725–4730 (2018).

33. Neve, P. Gene drive systems: do they have a place in agricultural weed management? *Pest Manag. Sci.* **74**, 2671–2679 (2018).

34. Barrett, L. G. et al. Gene drives in plants: opportunities and challenges for weed control and engineered resilience. *Proc. Biol. Sci.* **286**, 20191515 (2019).

35. Oye, K. A. et al. Regulating gene drives. *Science* **345**, 626–628 (2014).

36. Akbari, O. S. et al. Safeguarding gene drive experiments in the laboratory. *Science* **349**, 927–929 (2015).

37. James, S. L., O'Brochta, D. A., Randazzo, F. & Akbari, O. S. A gene drive is a gene drive: the debate over lumping or splitting definitions. *Nat. Commun.* **14**, 1749 (2023).

38. Metzloff, M. et al. Experimental demonstration of tethered gene drive systems for confined population modification or suppression. *BMC Biol.* **20**, 119 (2022).

39. Kim, Y. J., Zhang, D. & Jung, K. H. Molecular basis of pollen germination in cereals. *Trends Plant Sci.* **24**, 1126–1136 (2019).

40. Zhang, Q. et al. A calmodulin-binding protein from rice is essential to pollen development. *J. Plant Biol.* **55**, 8–14 (2012).

41. Fagard, M. & Vaucheret, H. (Trans) gene silencing in plants: how many mechanisms? *Annu. Rev. Plant Biol.* **51**, 167–194 (2000).

42. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

43. Muralla, R., Lloyd, J. & Meinke, D. Molecular foundations of reproductive lethality in *Arabidopsis thaliana*. *PLoS ONE* **6**, e28398 (2011).

44. Liu, H. et al. CRISPR-P 2.0: an improved CRISPR–Cas9 tool for genome editing in plants. *Mol. Plant* **10**, 530–532 (2017).

45. Xie, X. et al. CRISPR-GE: a convenient software toolkit for CRISPR-based genome editing. *Mol. Plant* **10**, 1246–1249 (2017).

46. Yoo, S. D., Cho, Y. H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).

47. Wu, F. H. et al. Tape-*Arabidopsis* sandwich – a simpler *Arabidopsis* protoplast isolation method. *Plant Methods* **5**, 16 (2009).

48. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

50. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

51. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom. Bioinform.* **3**, lqab019 (2021).

52. Schirmer, M., D'Amore, R., Ijaz, U. Z., Hall, N. & Quince, C. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*. **17**, 125 (2016).

53. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).

54. Liu, Y. G., Mitsukawa, N., Oosumi, T. & Whittier, R. F. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463 (1995).

55. Berardini, T. Z. et al. The *Arabidopsis* information resource: making and mining the 'gold standard' annotated reference plant genome. *Genesis* **53**, 474–485 (2015).

56. Li, D. et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).

57. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

58. Chen, T. et al. The genome sequence archive family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics* **19**, 578–583 (2021).

59. CNCB-NGDC Members and Partners.Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024.*Nucleic Acids Res*. **52**, D18–D32 (2024).

Author contributions

W.Q. and Y.L. designed the study. Y.L. performed the experiments. B.J. and Y.L. performed the computational analyses. Y.L., J.C. and W.Q. wrote the manuscript.

Competing interests

Y.L., B.J. and W.Q. have been granted a China invention patent (ZL202311247476.0) based on some results reported in this paper. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-024-01692-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-024-01692-1.

**Correspondence and requests for materials** should be addressed to Wenfeng Qian.

Peer review information *Nature Plants* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
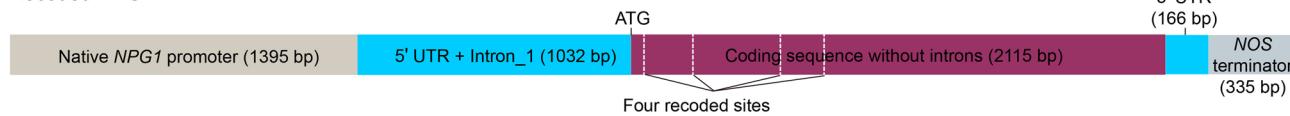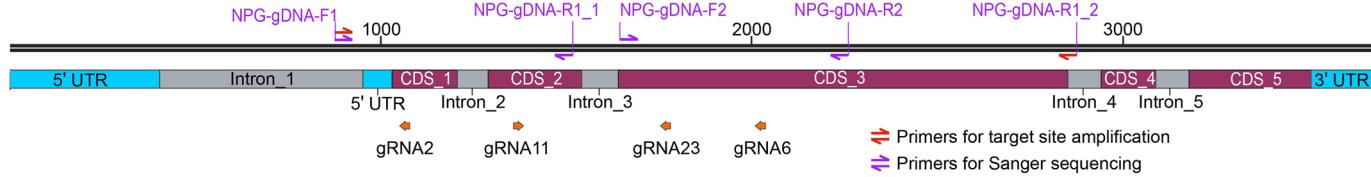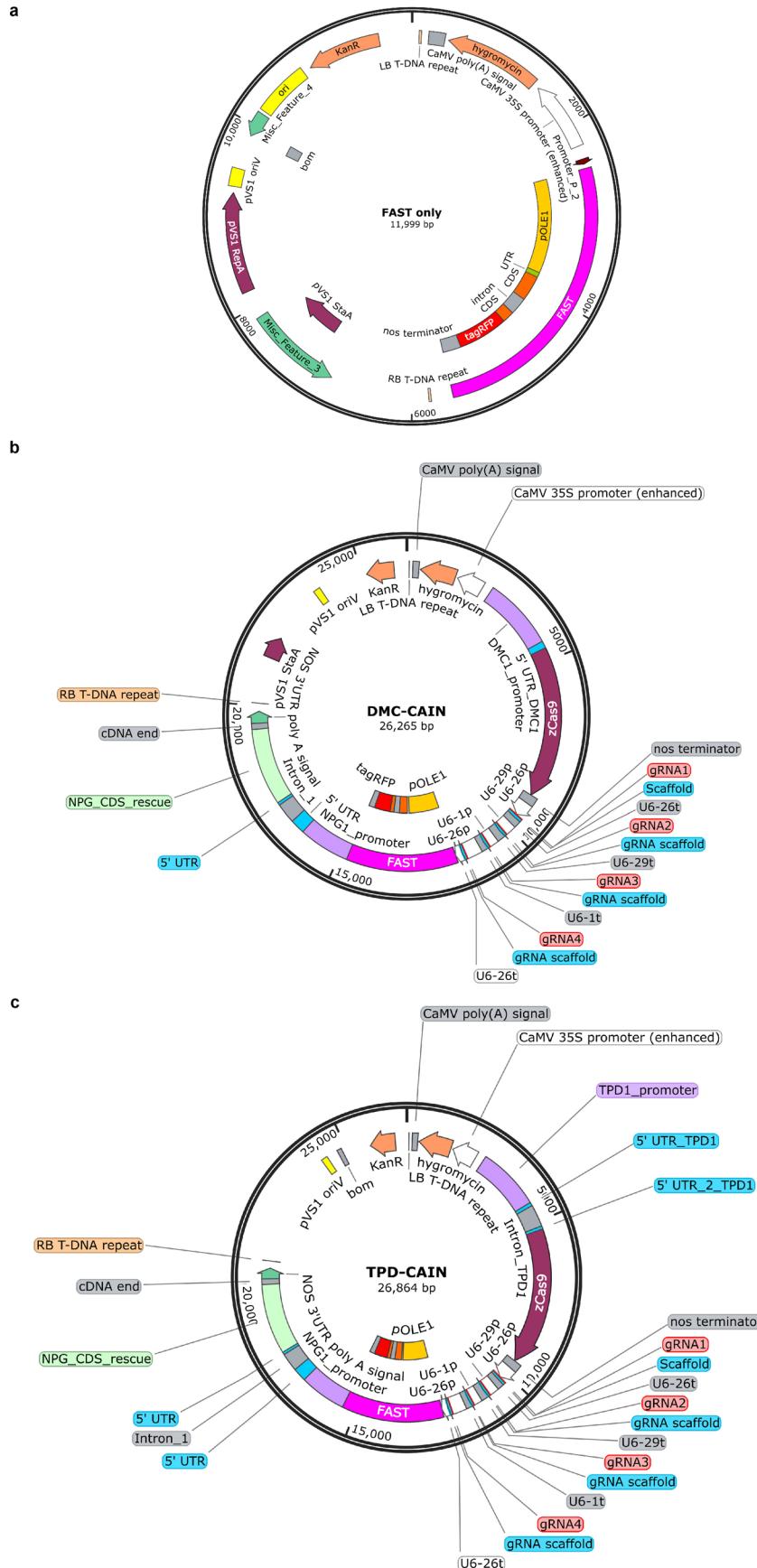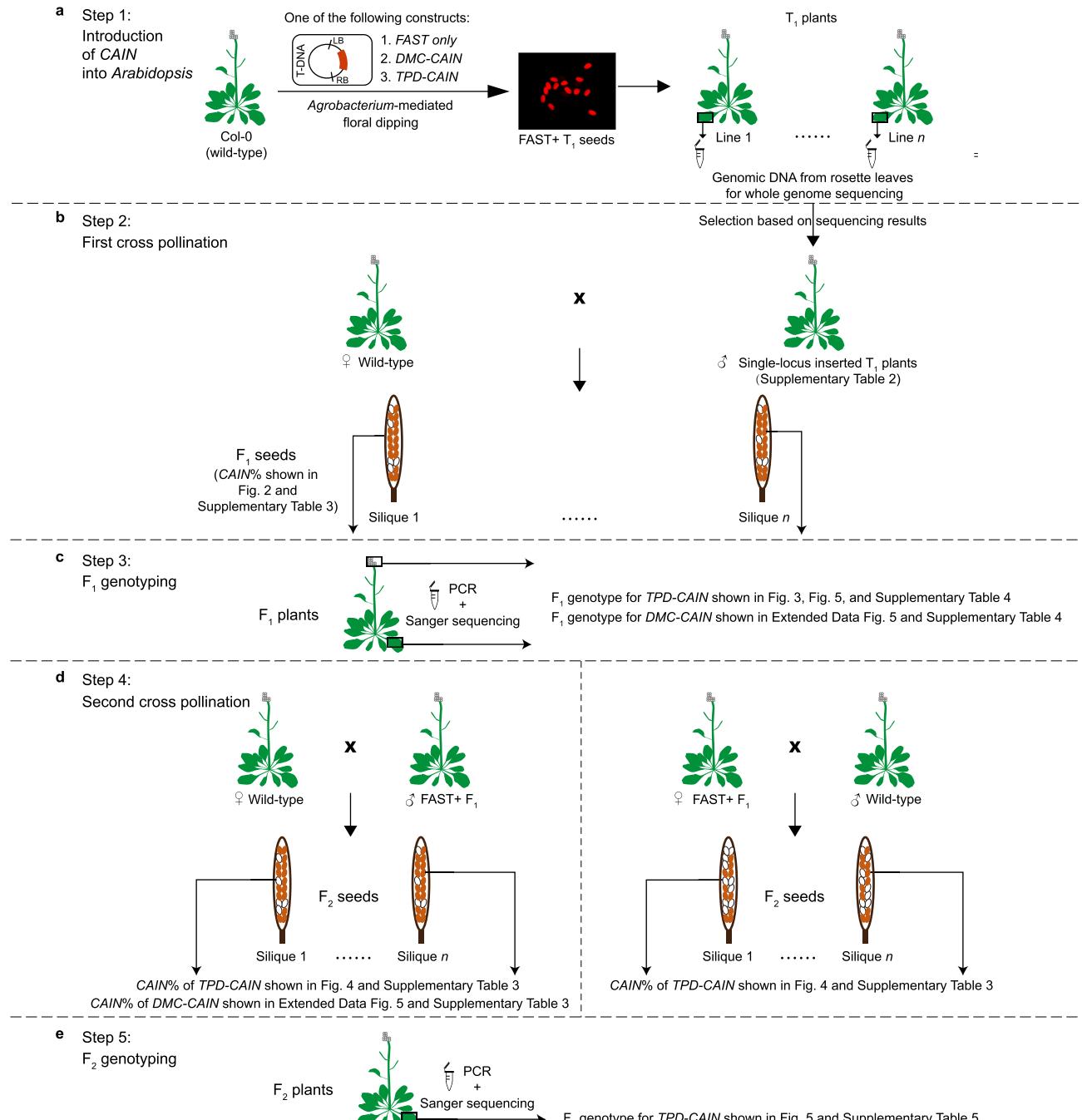
Extended Data Fig. 1 | Incorporation of four gRNAs into CAIN. a, Four gRNAs selected to be multiplexed into the *CAIN* constructs. Mutated nucleotides in the recoded version of *NPG1*, based on synonymous codons, are indicated by red squares. **b**, The structure of recoded *NPG1*. **c**, The position of four gRNA target sites on genomic *NPG1* allele. Amplification of genomic regions containing the four target sites was performed using the NPG-gDNA-F1 and NPG-gDNA-R1_2 primers (in red). Sanger sequencing of the target sites was performed using four primers (NPG-gDNA-F1, NPG-gDNA-F2, NPG-gDNA-R1_1, and NPG-gDNA-R2, in purple).

Extended Data Fig. 2 | Schematic maps of FAST only control construct and *CAIN* gene drives. Shown are schematic maps of the FAST only control construct (**a**), *DMC-CAIN* (**b**), and *TPD-CAIN* (**c**) gene drive constructs, illustrating their total length and main features.

Extended Data Fig. 3 | Experimental procedures overview. a, Step 1: Introduction of control construct (FAST only), *DMC-CAIN*, or *TPD-CAIN* gene drive construct into *Arabidopsis* Col-0 plants, indicated by red fluorescence marker (FAST +) in dry seeds. Single-locus insertion screening was performed using TAIL-PCR and whole genome resequencing. **b**, Step 2: Screened T₁ plants were used as the male parent to cross with wild-type female parent. Transmission rate of the CAIN gene drive (*CAIN%*) was calculated from the fraction of F₁ seeds exhibiting FAST marker. **c**, Step 3: Cultivation of F₁ seeds to obtain F₁ plants and genotyping (as described in Methods). **d**, Step 4: Crosses between F₁ plants of known genotypes (male or female parent) and wild-type plants, calculating *CAIN%* in F₂ seeds. **e**, Step 5: Cultivation of F₂ seeds to obtain F₂ plants and genotyping.
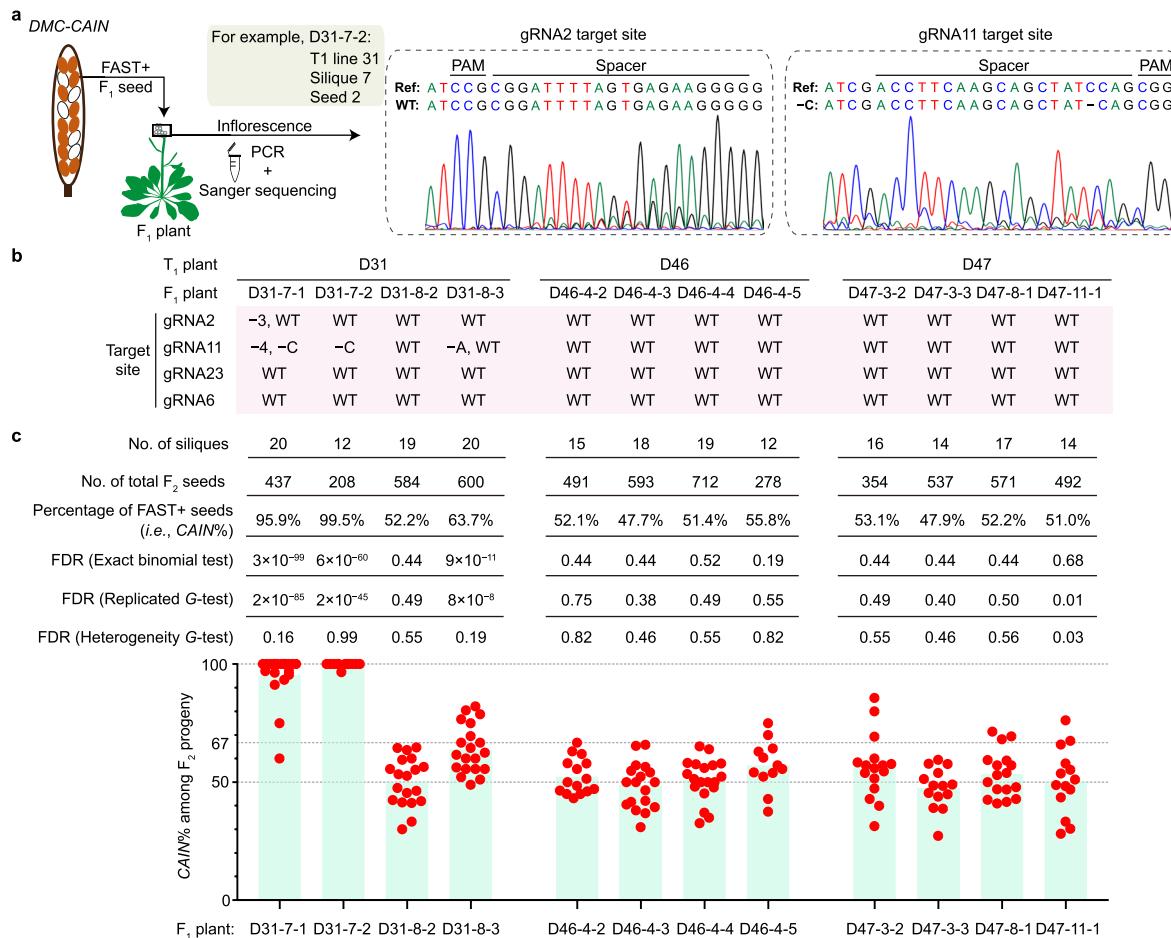
a

gRNA2 PAM | Spacer

```
Ref:    CCGCGG  A  TTTTAGTGAGAAGGGGGAAGATGAGATCGTCAGACAGCTT
-15*    CCGCGG  -  --------------GGGAAGATGAGATCGTCAGACAGCTT
-A      CCGCGG  -  TTTTAGTGAGAAGGGGGAAGATGAGATCGTCAGACAGCTT
+A      CCGCGG  AATTTTAGTGAGAAGGGGGAAGATGAGATCGTCAGACAGCTT
-41*    C-----  -  ---------------------------------AGCTT
+G      CCGCGGGA  TTTTAGTGAGAAGGGGGAAGATGAGATCGTCAGACAGCTT
A>C     CCGCGG  C  TTTTAGTGAGAAGGGGGAAGATGAGATCGTCAGACAGCTT
G>T     CCGCGT  A  TTTTAGTGAGAAGGGGGAAGATGAGATCGTCAGACAGCTT
```
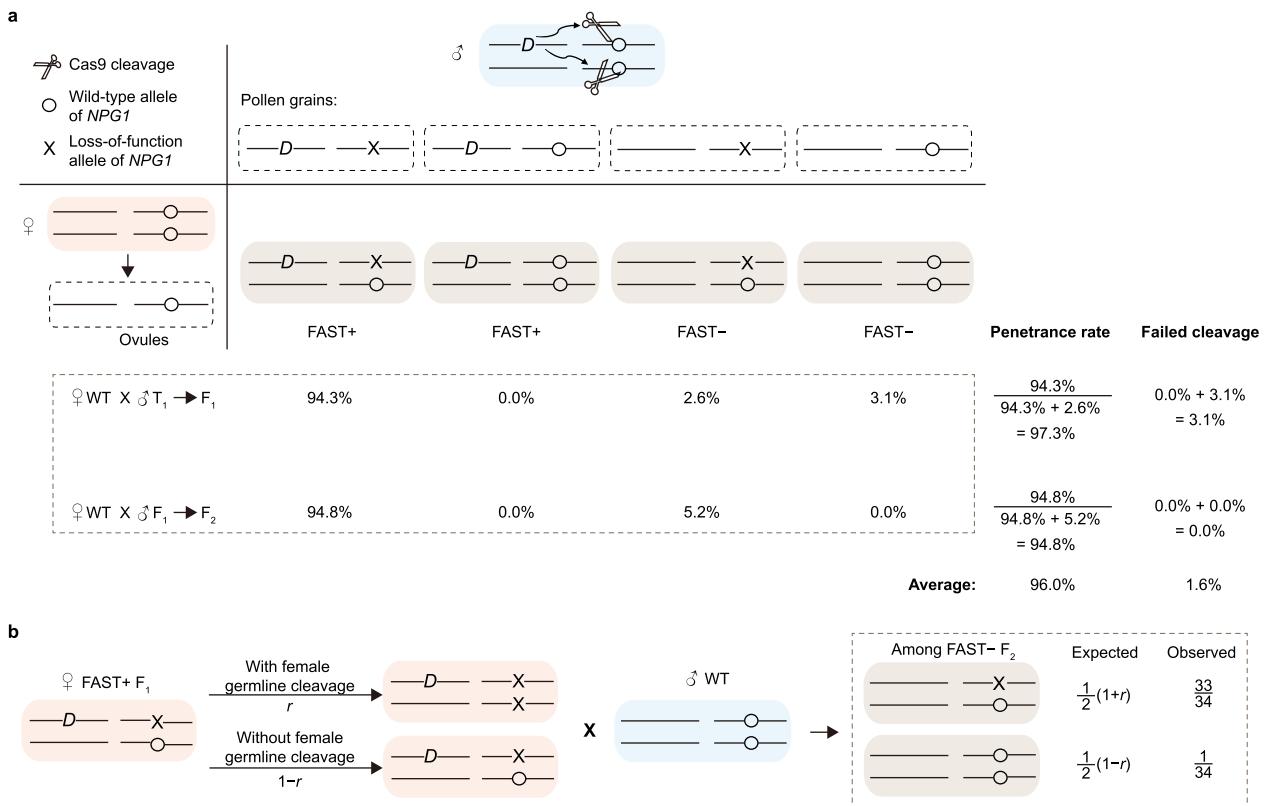
b

gRNA11 | Spacer | PAM

```
Ref:    ACCTTCAAGCAGCTA        TC  CAGCGGTTACAGGTTTCCGTGCCTCTT
-C      ACCTTCAAGCAGCTA        TC  -AGCGGTTACAGGTTTCCGTGCCTCTT
-8*     ACCTTCAAG------        --  CAGCGGTTACAGGTTTCCGTGCCTCTT
+T      ACCTTCAAGCAGCTA        TCTCAGCGGTTACAGGTTTCCGTGCCTCTT
-TC     ACCTTCAAGCAGCTA        --  CAGCGGTTACAGGTTTCCGTGCCTCTT
-18+8   ACCTTCAAGCAGCTACTAAGAAA--  ---------------CCGTGCCTCTT
+A      ACCTTCAAGCAGCTA        TCACAGCGGTTACAGGTTTCCGTGCCTCTT
+C      ACCTTCAAGCAGCTA        TCCCAGCGGTTACAGGTTTCCGTGCCTCTT
-3      ACCTTCAAGCAGCT-        --  CAGCGGTTACAGGTTTCCGTGCCTCTT
-4+A    ACCTTCAAGCAGC--        -A  CAGCGGTTACAGGTTTCCGTGCCTCTT
-5*     ACCTTCAAGCAG---        --  CAGCGGTTACAGGTTTCCGTGCCTCTT
-6      ACCTTCAAGCA----        --  CAGCGGTTACAGGTTTCCGTGCCTCTT
-25     ACCTTCAAGCAGCTA        --  --------------------TCTT
-37*    A-------------        --  --------------------CCTCTT
```

Extended Data Fig. 4 | Mutation types detected in FAST+ F₁ plants. This figure illustrates the locations of specific bases involved in insertions, deletions, and single nucleotide polymorphisms generated at the gRNA2 (**a**) and gRNA11 (**b**) target sites. * denotes one potential alignment result, as the underlined bases can be positioned on either side of the deletion.
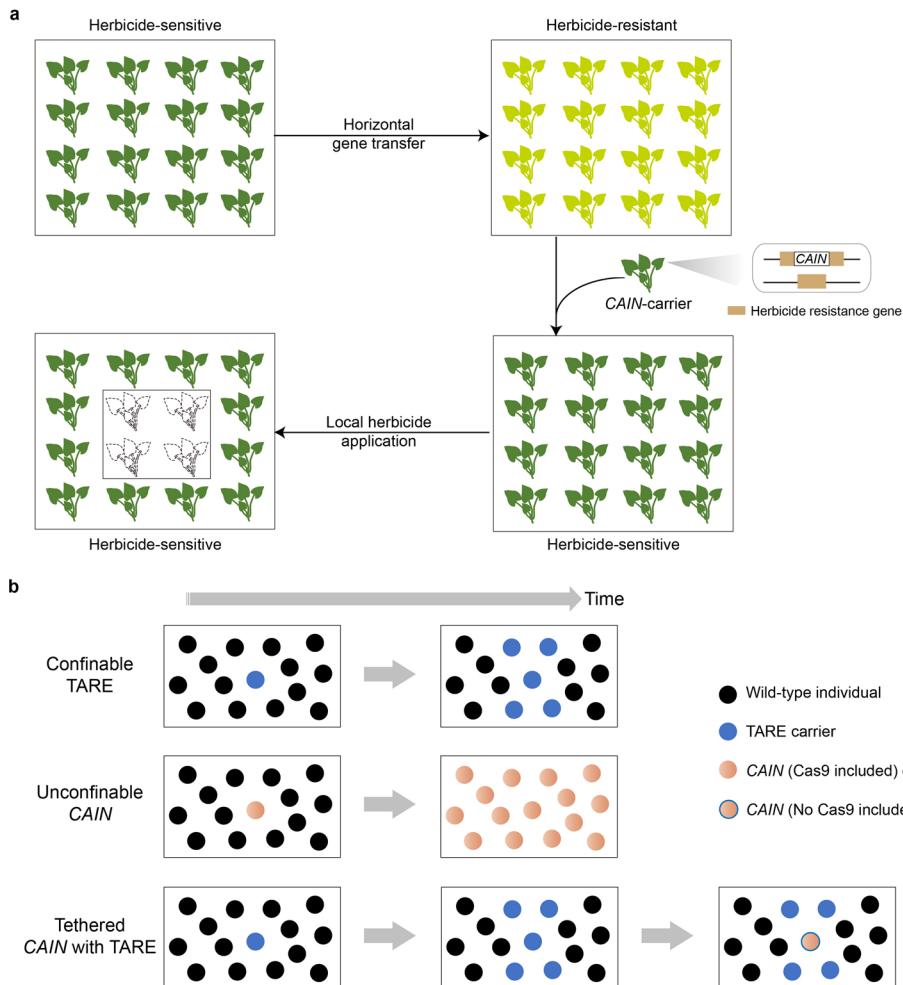
Extended Data Fig. 5 | Transmission rate of *DMC-CAIN* from the F1 to F2 generation. a, Diagram of genotyping in F_1 plants' inflorescence. **b**, Sanger sequencing-based genotyping results for 12 FAST + F_1 plants at four gRNA target sites. **c**, Bar plots illustrating the mean *DMC-CAIN* transmission rate estimated from individual siliques.

a

	FAST+	FAST+	FAST−	FAST−	Penetrance rate	Failed cleavage
♀ WT × ♂ T_1 → F_1	94.3%	0.0%	2.6%	3.1%	$\frac{94.3\%}{94.3\% + 2.6\%}$ = 97.3%	0.0% + 3.1% = 3.1%
♀ WT × ♂ F_1 → F_2	94.8%	0.0%	5.2%	0.0%	$\frac{94.8\%}{94.8\% + 5.2\%}$ = 94.8%	0.0% + 0.0% = 0.0%
				Average:	96.0%	1.6%

b

Among FAST− F_2 | Expected | Observed

$\frac{1}{2}(1+r)$ — $\frac{33}{34}$

$\frac{1}{2}(1-r)$ — $\frac{1}{34}$
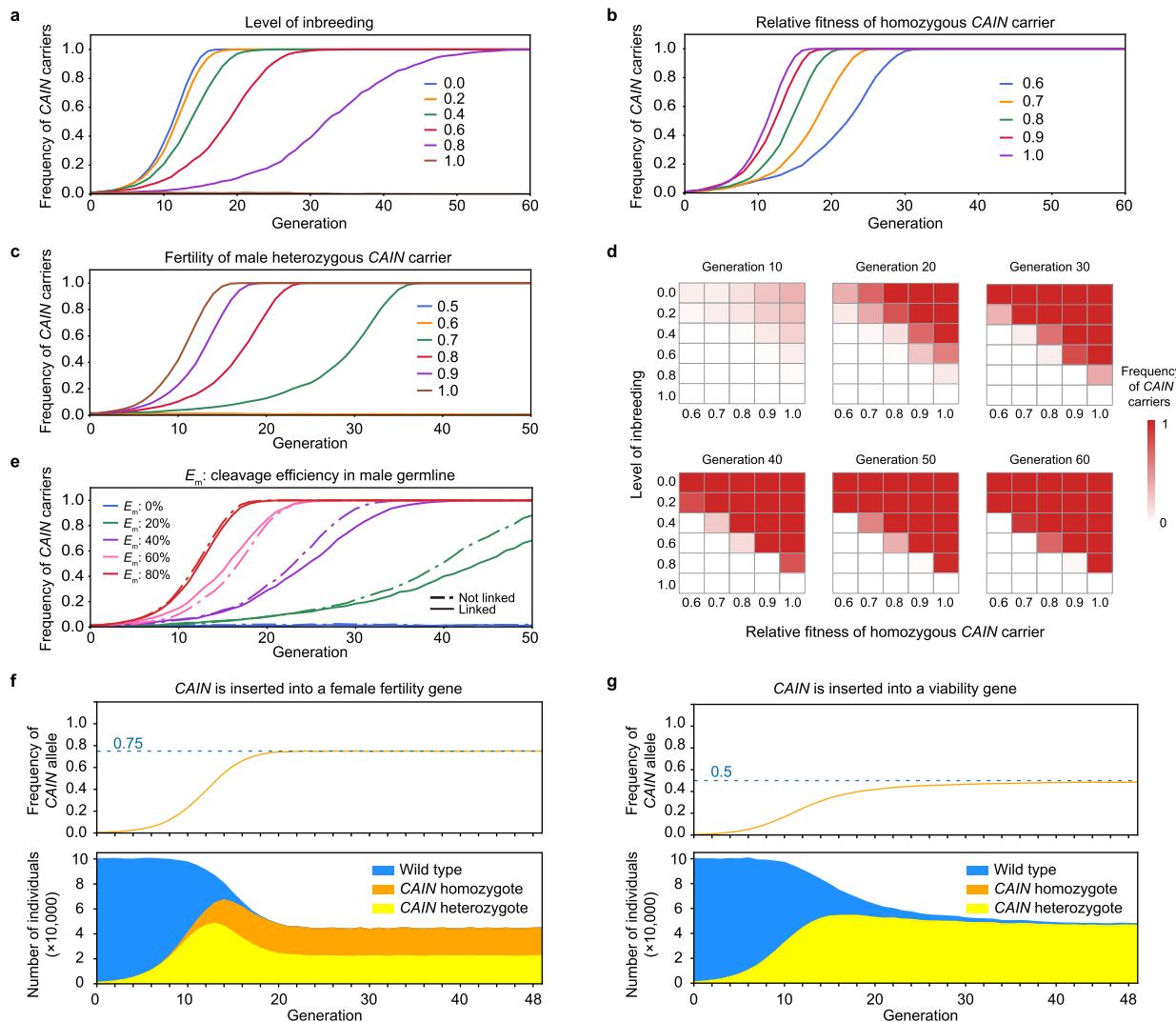
Extended Data Fig. 6 | Estimation of male germline cleavage efficiency, incomplete penetrance, and female germline cleavage efficiency of the *TPD-CAIN* system. a, Estimated male germline cleavage efficiency and penetrance rate in F_1 and F_2 progeny. Among the F_1 progeny, 94.3% (526/558, Supplementary Table 3) of the plants were FAST+ (*TPD-CAIN*/+), with all of them showing an *NPG1*⁻ genotype at the gRNA11 target site (Supplementary Table 4). A small percentage, 2.6% (5.7% × 5/11), were +/+; *NPG1*⁺/⁻ plants, and another 3.1% (5.7% × 6/11) were +/+; *NPG1*⁺/⁺ plants. When examining our F2 progeny, 94.8% (3868 out of 4080, Supplementary Table 3) were FAST+ (*TPD-CAIN*/+), all of which showed an *NPG1*⁻ genotype at the gRNA11 target site (Supplementary

Table 5). The remaining 5.2% F_2 plants were all +/+; *NPG1*⁺/⁻ (Supplementary Table 5). From these results, we estimated an average failed cleavage rate of 1.6%, which corresponds to a male germline cleavage efficiency of 98.4%. Additionally, we estimated an average penetrance rate of 96.0%. **b**, Estimation of female germline cleavage efficiency (*r*). We calculated the cleavage efficiency *r* using the genotypes of the *NPG1* locus at the gRNA11 target site in FAST− F_2 plants, assuming no further cleavage occurs in these plants. From an observed fraction of 1/34 +/+; *NPG1*⁺/⁺ plants (Supplementary Table 5), we estimated *r* to be 94.1%.

Extended Data Fig. 7 | Potential application and confinement of the *CAIN* system. a, Weed management application: inserting the *CAIN* system into a herbicide resistance gene of target weeds, and combining this with localized herbicide applications in the field, enables *CAIN* to facilitate spatially confined weed suppression. **b**, Confinement of *CAIN* spread via tethering with TARE. The TARE drive (or a similar variant) is confinable based on the initial release frequency of drive-carriers into the population (upper panel), in contrast to the *CAIN* system, which lacks such confinable characteristics (middle panel). To address its unconfinable nature, the *CAIN* (without *Cas9* included in the drive construct), when tethered to TARE, exhibits functionality contingent upon the presence of Cas9 provided by TARE. This arrangement effectively confines the spread of the *CAIN* system (bottom panel).

Extended Data Fig. 8 | Simulation analysis of factors influencing the spread dynamics of the *CAIN* system. a-c, Computational simulations depicting the impact of inbreeding (**a**), fitness cost (**b**), and the fertility of male heterozygous *CAIN* carriers (**c**) on the spread dynamics of the *CAIN* system. **d**, Computational simulation depicting the combined effect of inbreeding and fitness cost on the

spread dynamics of the *CAIN* system. **e**, Computational simulation depicting the spread dynamics of the *CAIN* system, both linked and unlinked with its target gene. **f**, The spread dynamics of the *CAIN* system inserted into a female fertility gene. **g**, The spread dynamics of the *CAIN* system inserted into a viability gene.

nature portfolio

Corresponding author(s): Wenfeng Qian

Last updated by author(s): Oct 10, 2023

Reporting Summary

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

Statistics

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

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *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	No software is used.
Data analysis	fastp (version 0.23.1), Burrow-Wheeler Aligner (version 0.7.17-r1188), SAMtools (version 1.13), Megahit (version 1.2.9), Blastn (version 2.9.0 +), R (version 4.2.3), RStudio (version 2023.03.0+386), R package DescTools (version 0.99.49), Perl (version 5.16.3), Python (version 3.10), CRISPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/), CRISPR-GE (http://skl.scau.edu.cn/home/). All relevant in-house scripts are available at GitHub (https://github.com/QianLabWebsite/GeneDrive).

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Illumina sequencing data have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender (identity/presentation), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Not applicable.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable.
Population characteristics	Not applicable.
Recruitment	Not applicable.
Ethics oversight	Not applicable.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	A basic power analysis of binomial test was performed. To acquire the power of 90%, we should analyze at least 93 seeds for a 17% increase from 50% transmission rate.
Data exclusions	No data were excluded from the analyses.
Replication	Biological replicates were performed as described in the main text. To assess the transmission rate of the gene drive in F1 progeny, three and four T1 plants were used as male parents for DMC-CAIN and TPD-CAIN, respectively. For F1 reciprocal crosses aimed at producing F2 progeny, we utilized at least eight F1 plants for both cross directions. Progeny were individually collected from each silique. The experimental findings were supported by all biological replicates.
Randomization	Plants were screened and separated based on genotype (indicated by the presence of red fluorescence in dry seed stage) and then individuals were randomly selected from this pool for following hybridization or target site genotyping as described.
Blinding	The investigators were blinded to seed/plant selection within each independent line during cross-pollination.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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