nature plants

LETTERS

https://doi.org/10.1038/s41477-020-0663-x

Check for updates

CRISPR–Cas9-mediated induction of heritable chromosomal translocations in *Arabidopsis*

Natalja Beying[1], Carla Schmidt[1], Michael Pacher[1], Andreas Houben[2] and Holger Puchta[1]✉

Clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein (Cas) technology has been applied in plant breeding mainly on genes for improving single or multiple traits[1–4]. Here we show that this technology can also be used to restructure plant chromosomes. Using the Cas9 nuclease from *Staphylococcus aureus*[5], we were able to induce reciprocal translocations in the Mbp range between heterologous chromosomes in *Arabidopsis thaliana*. Of note, translocation frequency was about five times more efficient in the absence of the classical non-homologous end-joining pathway. Using egg-cell-specific expression of the Cas9 nuclease and consecutive bulk screening, we were able to isolate heritable events and establish lines homozygous for the translocation, reaching frequencies up to 2.5% for individual lines. Using molecular and cytological analysis, we confirmed that the chromosome-arm exchanges we obtained between chromosomes 1 and 2 and between chromosomes 1 and 5 of *Arabidopsis* were conservative and reciprocal. The induction of chromosomal translocations enables mimicking of genome evolution or modification of chromosomes in a directed manner, fixing or breaking genetic linkages between traits on different chromosomes. Controlled restructuring of plant genomes has the potential to transform plant breeding.

Given the challenges of feeding the rapidly growing human population and the effects of climate change on agriculture, there is increasing demand for new crop varieties. As conventional breeding is reaching its limits, engineering crops for desirable traits using genome editing tools is becoming a major focus[6]. The application of the CRISPR–Cas system for targeted induction of site-specific double strand breaks (DSBs) has enabled use of gene editing in plants both for basic research and for the generation and improvement of agricultural traits[7].

In multicellular eukaryotes including plants, repair of DSBs is mediated mainly by two pathways, non-homologous end-joining (NHEJ) and homologous recombination[8]. Repair via error-prone NHEJ is often associated with loss of sequence information at the break site, whereas homologous recombination leads mostly to error-free repair[9]. In plants, NHEJ is the prevalent repair pathway in somatic tissue. NHEJ can be further subdivided into the classical NHEJ (cNHEJ) and alternative NHEJ (aNHEJ) pathways[10]. In the case of cNHEJ, broken ends are directly religated, sometimes resulting in small insertions or deletions (indels) at the break site. aNHEJ utilizes microhomologies close to the break sites and is dependent on polymerase theta, leading to deletions of the sequence information between the microhomologies partly associated with insertions[11,12].

The induction of several DSBs at a time can lead to complex rearrangements in a genome by joining unrelated break ends by NHEJ. Thus, if two DSBs are induced on the same chromosome, deletions or inversions can be achieved[13,14]. The simultaneous induction of two DSBs on heterologous chromosomes may lead to the formation of reciprocal translocations[15]. In mammals, translocations are associated with various genetic diseases and cancer[16–18]. In plants, translocations have an important role in trait diversity, speciation and genome evolution[19,20]. Such chromosomal rearrangements can lead to genetic isolation between populations through the suppression of recombination in heterozygotes. For breeders, stabilization of trait linkages in elite cultivars and breaking linkage drags can be advantageous. A technology for the controlled induction of heritable chromosomal translocations would be a possible solution for at least some of these challenges.

For translocation formation, we used the Cas9 nuclease from *Staphylococcus aureus* (SaCas9)[5] to induce two DSBs (TL1–2) on chromosome 1 (Chr1) and Chr2 of *A. thaliana*. The target sites of the Cas9 nuclease were designed to cut in intergenic regions, 0.5 Mbp from the end of the long chromosome arms of Chr1 and Chr2 (Fig. 1a). To determine the translocation frequency in somatic cells, we transformed the CRISPR constructs in Col-0 plants by *Agrobacterium*-mediated floral dipping and selected primary transformants after two weeks of growth. For each of the 6 biological replicates analysed, 100 plants were pooled for extraction of genomic DNA. Quantitative measurement of the translocation frequency was conducted using digital droplet PCR (ddPCR) with site-specific primers and probes for the newly formed junctions in a duplex reaction with a normalization control (Fig. 1b). For this amplification control, a sequence located proximal to the junction site was used to normalize the amount of translocation junction detected to the amount of genomes analysed. We detected translocation frequencies of around 0.01% at both junctions for TL1–2, indicating that reciprocal chromosome-arm exchanges are indeed achievable by the use of CRISPR–Cas9.

To determine which repair pathway is involved in the formation of a translocation in somatic plant cells, we checked the sequence pattern at the junction sites by next-generation sequencing (NGS) (Fig. 1d). In the Col-0 wild-type background, around 60% of events showed error-free ligation. The remaining events were deletions, mostly small in size (see also Extended Data Fig. 1b). In short, in most cases, no nucleotides were lost, most probably because the newly ligated chimeric junctions—in contrast to the original cut sites—could no longer be recognized by the Cas9 nuclease. This indicates that it is mainly the cNHEJ pathway that is responsible for the formation of chromosomal translocations. Therefore, we also analysed translocation formation in a line in which a key enzyme of cNHEJ, KU70, is knocked out. The KU70–KU80 heterodimer is essential for cNHEJ activity in protecting the DSB ends[21]. Knockout of KU70 increased the translocation frequency by approximately

[1]Botanical Institute—Molecular Biology and Biochemistry of Plants, Karlsruhe Institute of Technology, Karlsruhe, Germany. [2]Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, Seeland, Germany. ✉e-mail: holger.puchta@kit.edu

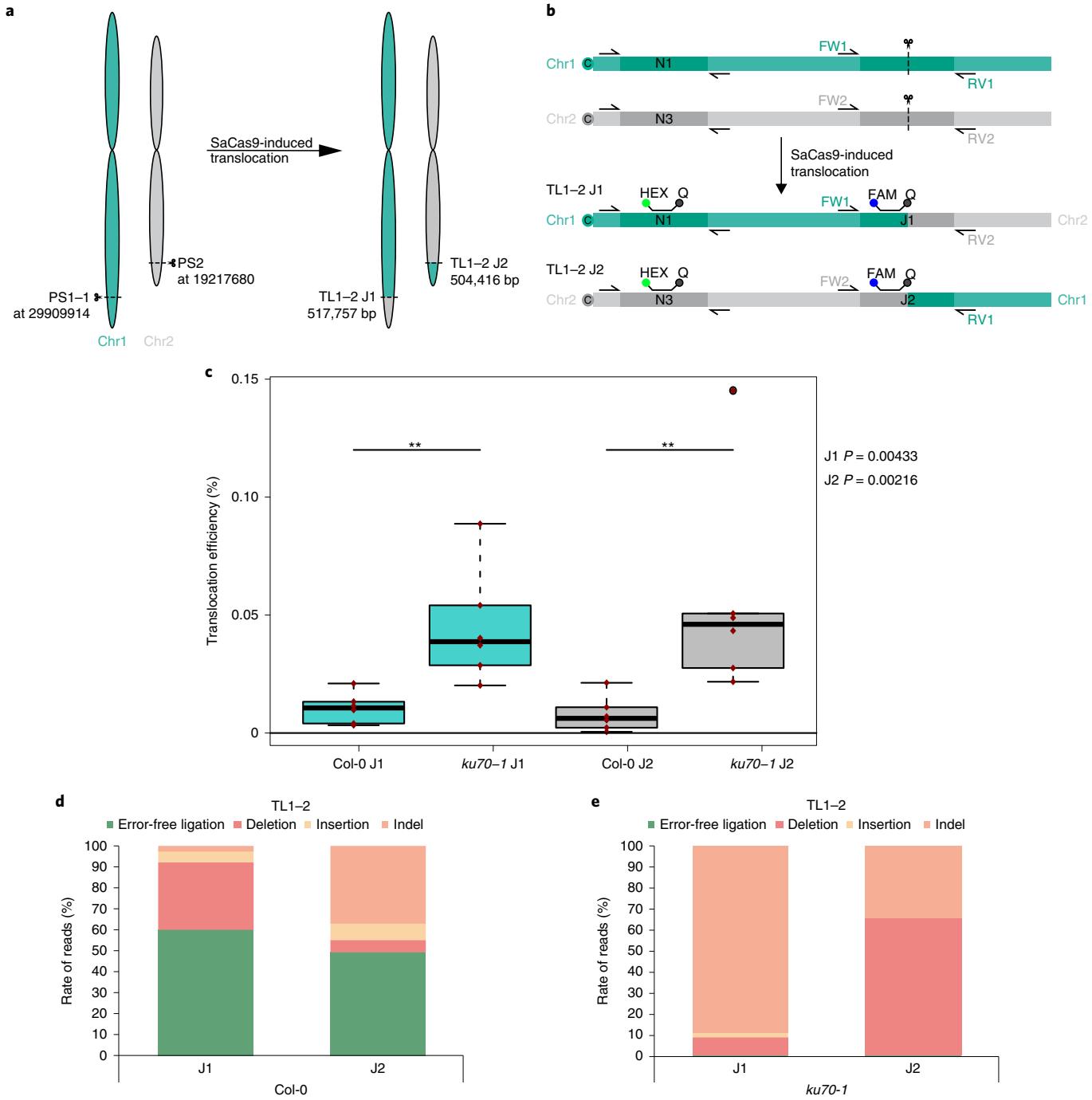


Fig. 1 | Induction of translocations between Chr1 and Chr2 in wild type and the *ku70-1* mutant by SaCas9. a, Using the multiplexed SaCas9 nuclease, two DSBs were induced on the long arms of Chr1 and Chr2 in *Arabidopsis*. Induction of these breaks may lead to a reciprocal chromosome-arm exchange of fragments of about 0.5 Mbp. **b**, Scheme of PCR analysis for determination of the sequence of the newly formed translocation (TL) junctions and the droplet digital PCR (ddPCR) assay for quantification. For ddPCR, both newly formed junctions (J1 and J2) were detected using site-specific primers (arrows) in combination with amplicon-specific probes. To set the detected amount of junctions relative to the number of genomes analysed, a duplex assay was performed with a normalization control located on the centromere-carrying chromosome. FAM, 6-fluorescein amidite; HEX, hexachloro-fluorescein; quencher Q, BHQ-1. **c**, The amount of translocation junctions relative to the genome number. In the wild type, similar amounts of J1 and J2 were detected. In *ku70-1* mutants, there was a significant increase of frequencies of both translocation junctions. For each box plot, six independent biological replicates with 100 T1 plants each were analysed ($n = 6$). In box plots, the middle line shows the median, box edges represent first and third quartiles, and error bars show s.d.; *P* values were calculated using the Mann–Whitney *U*-test ($U_{J1} = 1$, $U_{J2} = 0$; $^{**}P < 0.01$, two-tailed). **d**,**e**, Deep sequencing of the translocation junctions of the wild type (**d**) and the *ku70-1* mutant (**e**). Analysis revealed that around 60% of both junctions were ligated in an error-free manner in the wild-type background. In the *ku70* mutant, nearly all reads of the junctions analysed contained mutations, mainly deletions or indels.

fivefold, to nearly 0.05% (Fig. 1c). Sequence analysis revealed that most junctions contained deletions or indels in combination with microhomologies (Fig. 1d and Extended Data Fig. 1b–d). This demonstrates that aNHEJ is able to form translocations efficiently in the absence of cNHEJ. This result is reminiscent of a similar finding reported for intrachromosomal deletions and inversions[14]. Evidently, the KU70–KU80 heterodimer is required both for protection of the individual ends of a DSB from end resection and to keep the two ends linked together during the repair process[22]. In the absence of KU70, the probability that two previously unlinked ends meet and are rejoined by aNHEJ increases.

The tissue-specific EC1.1–EC1.2 promoter has previously been applied efficiently for Cas9 expression to create heritable mutations, deletions, inversions and gene-targeting events in *Arabidopsis*[23–26]. We therefore transformed wild-type Col-0 plants and the *ku70-1* mutant line with the same CRISPR construct but with egg-cell-specific Cas9 expression for heritable introduction of the translocation. Following the schematic in Fig. 2a, primary transformants were isolated by selection and subsequently selfed. The resulting T2 lines were bulk screened by PCR-based genotyping. This was done by analysing two pools of 40 T2 plants from each primary transformant. For Col-0, positive signals were detected for 3 out of 40 T2 lines for TL1–2 (Fig. 2f). In the *ku70* mutant line, 4 out of 20 T2 lines tested positive for the reciprocal TL1–2. Genomic DNA was extracted from each individual plant of the corresponding positive pool and screened for both translocation junctions. In each positive pool, we could confirm at least one plant carrying a reciprocal translocation. Thus, TL1–2 frequencies of up to 2.5% in individual T2 lines for the wild type and 3.75% for the *ku70* mutant were obtained. Overall, we were able to isolate four individual plants in the Col-0 background and eight individual plants in the *ku70-1* background exhibiting TL1–2. These results correspond to one out of 800 and one out of 200 screened plants with a heritable translocation, respectively, in line with the ddPCR analysis.

All four identified Col-0 plants and two *ku70* mutants carrying the translocation were selfed and PCR was used to analyse their progeny for segregation of the translocation junctions. The T2 plants should have both wild-type and chimeric copies of the respective chromosomes in their diploid chromosome set (Fig. 2b). Thus, half of their gametes should be genetically unbalanced, with two copies of one translocated chromosome arm and none of the other. As their genetic information is incomplete, these gametes might not be viable (Fig. 2c). Indeed, instead of the formally expected genotypes (Fig. 2d), we observed only three genotypes (homozygous wild type, heterozygous translocation and homozygous translocation; Fig. 2e) for the analysed Col-0 lines. Instead of the expected 1 out of 16 diploid zygotes being homozygous for the translocation (Fig. 2d), 20–30% of plants in the next generation had this genotype. This can easily be explained if we assume that only genetically balanced gametes can contribute to the germline. In such a situation, we would expect a 1:2:1 segregation pattern between homozygous wild-type, heterozygous translocation and homozygous translocation plants, which is fully in line with the results obtained (Fig. 2e). The same holds true for the *ku70* mutants: 17.5% to 30% of the progeny were homozygous for the translocation, with the same segregation pattern as in the wild-type background. Consequently, due to the selection for complete chromosome sets during sexual reproduction, we were able to obtain plants homozygous for the reciprocal translocation at much higher rates than formally expected.

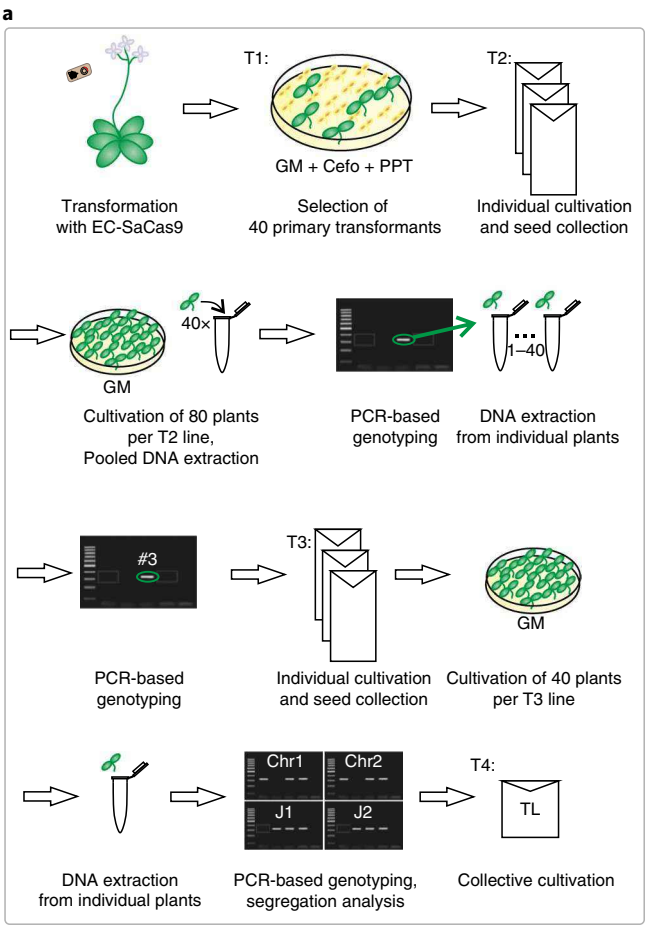
All plants identified as homozygous for the translocations were propagated and analysed to characterize the molecular nature of the translocations. Sequencing of the junction sites (Fig. 3a) confirmed a conservative and reciprocal chromosome-arm exchange. Perfect ligation without any sequence change at both junctions (J1 and J2) was observed in three (no. 6, 30 and 38) out of the four Col-0 lines carrying TL1–2. The fourth line (no. 24) contained a 44-base pair (bp) deletion at J1, but a perfectly ligated J2. As expected, the *ku70* mutant lines all had larger indels, the signature of aNHEJ. To demonstrate that no sequence information was lost, for every 100 kb on the translocated chromosome parts, a representative amplicon of around 2.5 kb was amplified for the homozygous lines no. 24 and 30 carrying the deletion at J1 and the perfectly ligated junction sites for TL1–2, respectively (Extended Data Fig. 2a). Moreover, both lines did not differ in phenotype and fertility from wild type (Extended Data Fig. 2b–d).

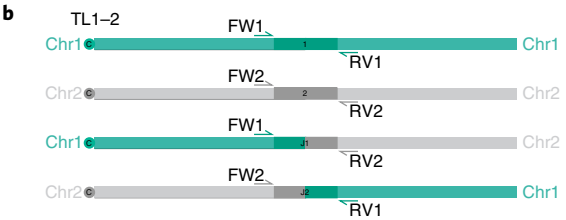
By tagging the chromosome arms at both ends of the break sites with differentially labelled DNA of predefined bacterial artificial chromosomes (BACs), we were able to show the translocation on a cytological level (Fig. 3a), providing further evidence that the chromosomes indeed carry the respective translocated chromosome arms. Whereas wild-type mitotic and naturally extended pachytene chromosomes displayed separate Chr1- and Chr2-specific fluorescence in situ hybridization (FISH) signals, translocated chromosomes of TL1–2 revealed combined FISH signals.

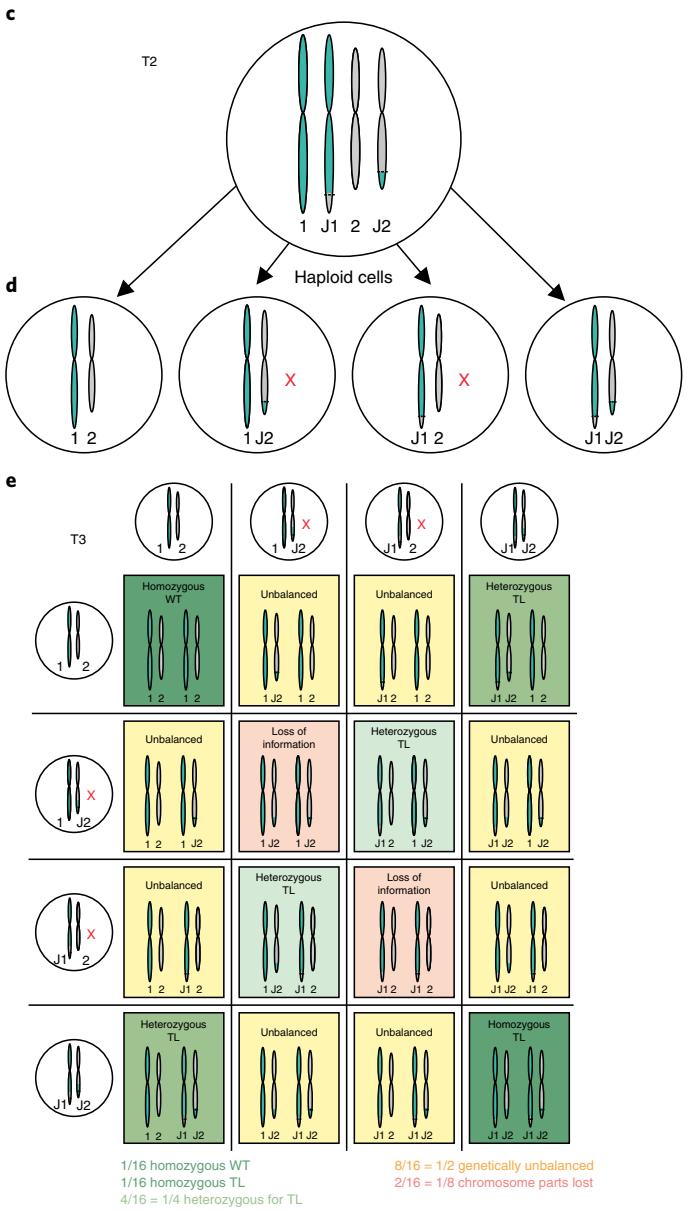
Fig. 2 | Establishment of lines homozygous for the translocation. a, Flow chart of the protocol for CRISPR–Cas9-mediated induction of translocation in *Arabidopsis*. After *Agrobacterium*-mediated transformation with SaCas9 under egg-cell (EC)-specific expression control, primary transformants were selected. Translocation events were induced in the egg cells of these T1 plants and pooled progeny of the respective plants were screened for the newly formed junctions. Individual plants in positive pools were then checked for recombinants by PCR-based screening. Plants carrying the reciprocal translocation were then propagated and plants homozygous for the translocation could be identified in the next generation. GM, germination medium; cefo, cefotaxime; PPT, phosphinotricin. **b**, For PCR-based screening, two primer combinations were designed to amplify the regions over the two induced DSBs. The translocation junction TL1–2 J1 can be detected by PCR with the forward primer on Chr1 (FW1) and the reverse primer on Chr2 (RV2). Conversely, TL1–2 J2 can be detected by PCR with the forward primer on Chr2 (FW2) and the reverse primer on Chr1 (RV1). **c**–**e**, Schematic segregation pattern of a reciprocal translocation in its hemizygous state. Chromosome sets of T2 plants harbour both wild-type (1 and 2) and chimeric (J1 and J2) copies of the respective chromosomes (**c**); their gametes therefore carry either the wild-type haploid chromosome set, the set harbouring the translocation, or two sets of chromosomes lacking part of the genetic information (marked with a red cross; **d**); in the T3 generation, 16 genotypic combinations can theoretically arise if all four gametes are transferred to the next generation without bias (**e**). However, half (8) of the fertilization events would result in unbalanced chromosome sets (yellow), two cells would lose genetic information (red) and only six combinations (green) would account for a balanced diploid progeny. Out of the six balanced genotypes, four are heterozygous for the translocation, meaning they carry both chromosomes in the original and the restructured configuration. Two of the genotypes would arise from unbalanced (light green) and two would from balanced gametes (dark green) combinations. Thus, if the unbalanced gametes were not able to contribute to the progeny, only 4 viable zygotes out of 16 (1 homozygous wild type, 2 heterozygous translocation and 1 homozygous translocation) would putatively be produced, resulting in a quasi-Mendelian segregation (1:2:1) of the translocated chromosomes. WT, wild type. **f**, Translocation frequencies and segregation pattern detected during line establishment. In T2 analysis, for all transformed lines, plants positive for both junction sites were detected, accounting for translocation frequencies of 0.125% and 0.5% for Col-0 and *ku70-1*, respectively. In the resulting T3 generation, 40 plants for each T3 line were screened and—as no unbalanced chromosome sets were detectable—a shift towards balanced progeny was observed. The χ^2-test was used to test for quasi-Mendelian segregation of the translocation junctions.

To demonstrate that the induction of heritable translocations was not loci dependent, we also induced two DSBs simultaneously at about 1 Mbp distance from the telomeres of the long arms of Chr1 and Chr5 (TL1–5; Fig. 4a) and screened for translocation formation. As the *ku70* mutant line, despite its higher induction efficiency, is genetically unstable if propagated over generations[27], we chose to transform the CRISPR constructs only in the Col-0 background. By screening 40 T2 pools, we were able to obtain a plant with a reciprocal translocation (Fig. 4b). The somewhat lower induction frequency of TL1–5 (0.03%) compared with that of TL1–2 (0.1%) is explained by the low cutting efficiency of the SaCas9 with protospacer 3 on Chr5 (Extended Data Fig. 1a). By selfing the positive T2 plant, we obtained progeny carrying the translocation homozygously at a rate of 10% (Fig. 4b). Our analysis of the molecular structure of the junction revealed small deletions of 1 bp in J1 and 12 bp in J2 (Fig. 4c). Using oligonucleotide-based painting probes, we demonstrated that the chromosome-arm exchange was conservative and reciprocal, as expected (Fig. 4d).

In summary, we have established a method to generate heritable targeted translocations (Fig. 2a). This demonstrates that we are able

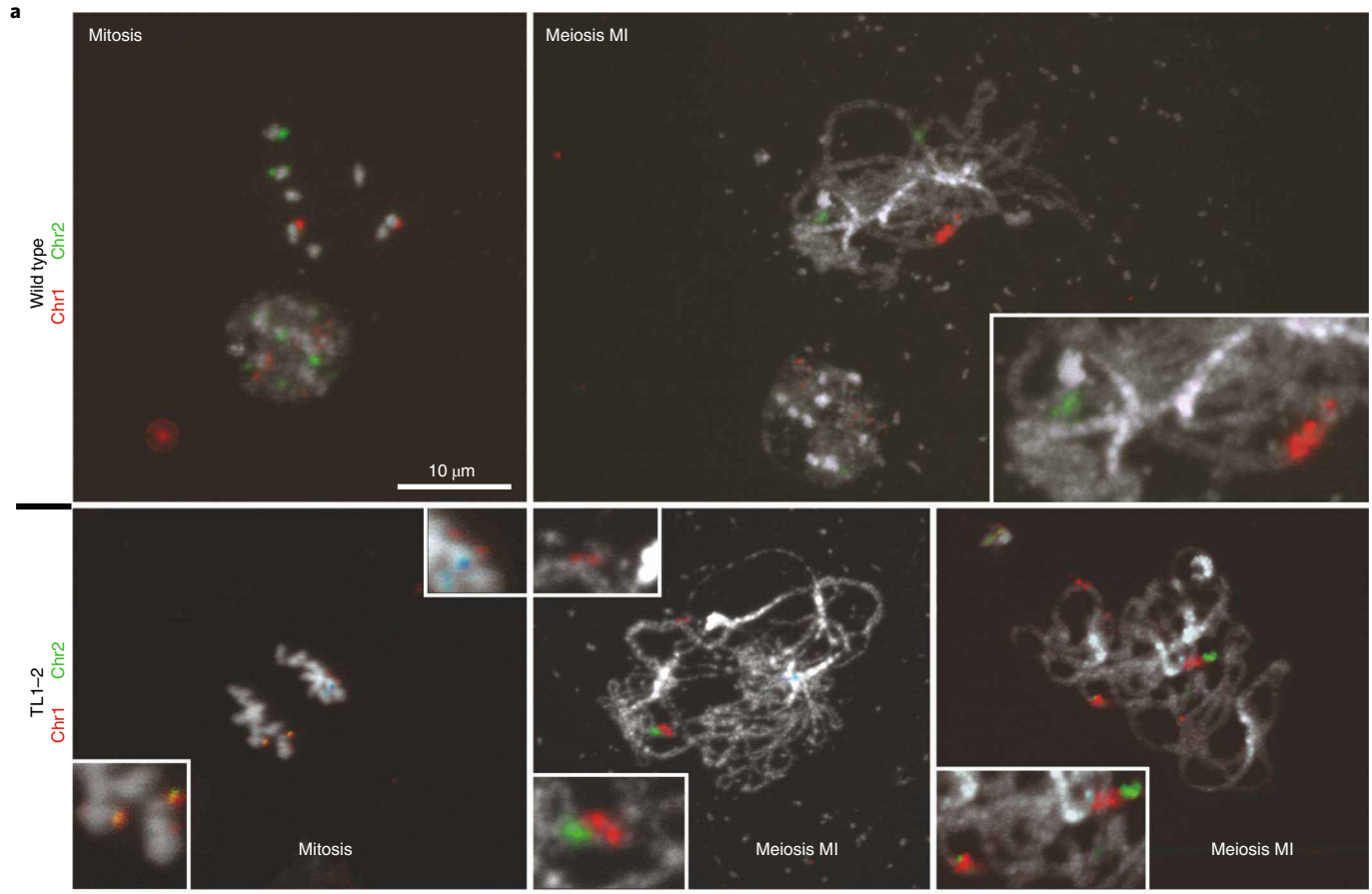






f

T1 line	Positive T2 lines out of lines analysed	Positive individual plants	T2 translocation frequencies	T3 line	Plants tested	Homozygous WT	Heterozygous TL	Homozygous TL	Unbalanced genotype	χ^2	Mendelian (1:2:1) segregation
Col-0::EC-TL1–2	3/40	4	0.125%	NBE657 no. 24	40	9	23	8	0	0.950	Yes
				NBE658 no. 6	40	10	18	12	0	0.600	Yes
				NBE669 no. 30	40	8	21	1	0	0.550	Yes
				NBE669 no. 38	40	12	20	8	0	0.800	Yes
ku70-1::EC-TL1–2	4/20	8	0.5%	NBE660 no. 11	40	13	15	12	0	2.550	Yes
				NBE715 no. 9	40	8	25	7	0	2.550	Yes



b

	TL1–2 J1	atatgcgagttatcagcacgcacctcaaatctgtaggtttaaccccctttggttattaaggtgtatggtttgattgcgaaagcaactcattccgaggtt	
Col-0	NBE657 no. 24	ATATGCGAGTTAT----------TAAGGTGTATGGTTTGATTGCGAAAGCAACTCATTCCGAGGTT	Δ44 bp
Col-0	NBE658 no. 6	ATATGCGAGTTATCAGCACGCACCTCAAATCTGTAGGTTTAACCCCCTTTTGGTTATTAAGGTGTATGGTTTGATTGCGAAAGCAACTCATTCCGAGGTT	Perfect
Col-0	NBE669 no. 30	ATATGCGAGTTATCAGCACGCACCTCAAATCTGTAGGTTTAACCCCCTTTTGGTTATTAAGGTGTATGGTTTGATTGCGAAAGCAACTCATTCCGAGGTT	Perfect
Col-0	NBE669 no. 38	ATATGCGAGTTATCAGCACGCACCTCAAATCTGTAGGTTTAACCCCCTTTTGGTTATTAAGGTGTATGGTTTGATTGCGAAAGCAACTCATTCCGAGGTT	Perfect
ku70-1	NBE660 no. 11	ATATGCGAGTTATCAGCAC----------TA----------TATTAAGGTGTATGGTTTGATTGCGAAAGCAACTCATTCCGAGGTT	Δ35 bp/+2 bp
ku70-1	NBE715 no. 9	ATATGCGAGTTATCAGCACGCACCTCAAATCTGTAGGTTT--GAGTTTTTTTTGAACCTACAAGCACCAAATCAAGCA-AAAGCAACTCATTCCGAGGTT	Δ39 bp/+36 bp

c

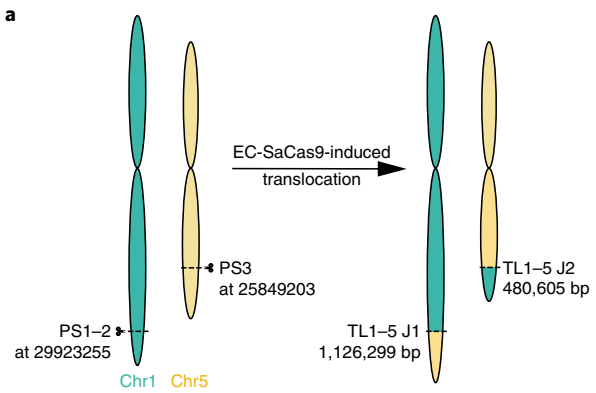
	TL1–2 J2	ggttgatggttcaaaaattaattagcttgatttggtttataaccctcccacgaaattagctatttttgtatagttcgaaaaagtaagcgcgtgatgtttt	
Col-0	NBE657 no. 24	GGTTGATGGTTCAAAAATTAATTAGCTTGATTTGGTTTATAACCCTCCCACGAAATTAGCTATTTTTGTATAGTTCGAAAAAGTAAGCGCGTGATGTTTT	Perfect
Col-0	NBE658 no. 6	GGTTGATGGTTCAAAAATTAATTAGCTTGATTTGGTTTATAACCCTCCCACGAAATTAGCTATTTTTGTATAGTTCGAAAAAGTAAGCGCGTGATGTTTT	Perfect
Col-0	NBE669 no. 30	GGTTGATGGTTCAAAAATTAATTAGCTTGATTTGGTTTATAACCCTCCCACGAAATTAGCTATTTTTGTATAGTTCGAAAAAGTAAGCGCGTGATGTTTT	Perfect
Col-0	NBE669 no. 38	GGTTGATGGTTCAAAAATTAATTAGCTTGATTTGGTTTATAACCCTCCCACGAAATTAGCTATTTTTGTATAGTTCGAAAAAGTAAGCGCGTGATGTTTT	Perfect
ku70-1	NBE660 no. 11	GGTTGATGGTT----------Δ158----------AAA----------	Δ158 bp/+3 bp
ku70-1	NBE715 no. 9	GGTTGATGGTTCAAAAATTAATTAGCTTGATTTGGTTTATA----------ATGAGTTTTTTTTTCACCCTC----------GTTTT	Δ54 bp/+21 bp

Fig. 3 | Analysis of lines carrying the translocation homozygously. a, Mitotic and naturally extended pachytene chromosomes of wild-type and TL1–2 plants after FISH using differently labelled DNA probes (BAC pools) specific for translocation-breakpoint regions of Chr1 (red) and Chr2 (green). Insets show further magnified chromosome regions. Experiments were repeated three times independently with similar results. MI, meiosis I. **b**,**c**, Sequencing results at the two junction sites J1 (**b**) and J2 (**c**) of the heritable translocation events. Three out of four T4 lines in the wild-type background showed error-free ligation at both junctions, and the other line harboured a small deletion at one of the junctions. Both analysed translocation lines on the *ku70-1* background harbour mutations characteristic of repair by aNHEJ at both junctions. PAM, protospacer adjacent motif.

to use CRISPR–Cas9 technology for both gene editing and chromosome engineering in plants. Similar systems could be established in crop plants for breeders to induce reciprocal translocations to fix or break genetic linkages. Moreover, it will also be possible to mimic chromosomal rearrangements as they occurred during plant genome evolution.

Methods

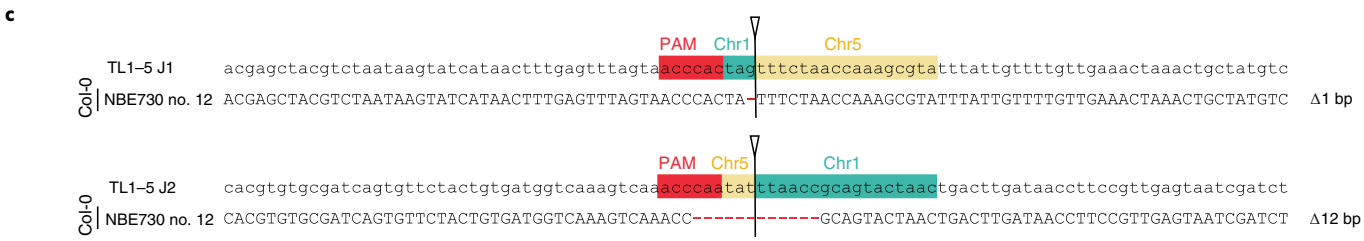
Construction of CRISPR–Cas9 vectors. T-DNA constructs used in this study are based on the Gateway compatible pEn-Sa-Chimera and pDe-SaCas9 or pDe-SaCas9-ECP plasmids, as previously described[3,24]. The destination vector pDe-SaCas9 expresses SaCas9 under the control of the PcUbi4-2 promoter together with the pea3A terminator for constitutive expression. In pDe-SaCas9-ECP, the SaCas9 is under the control of an egg-cell-specific promoter



b

T1 line	Positive T2 lines out of lines analysed	Positive individual plants	T2 translocation frequencies
Col-0::EC-TL1–5	1/40	1	0.031%

T3 line	Plants tested	Homozygous WT	Heterozygous TL	Homozygous TL	Unbalanced genotype
NBE730 no. 12	40	17	19	4	0



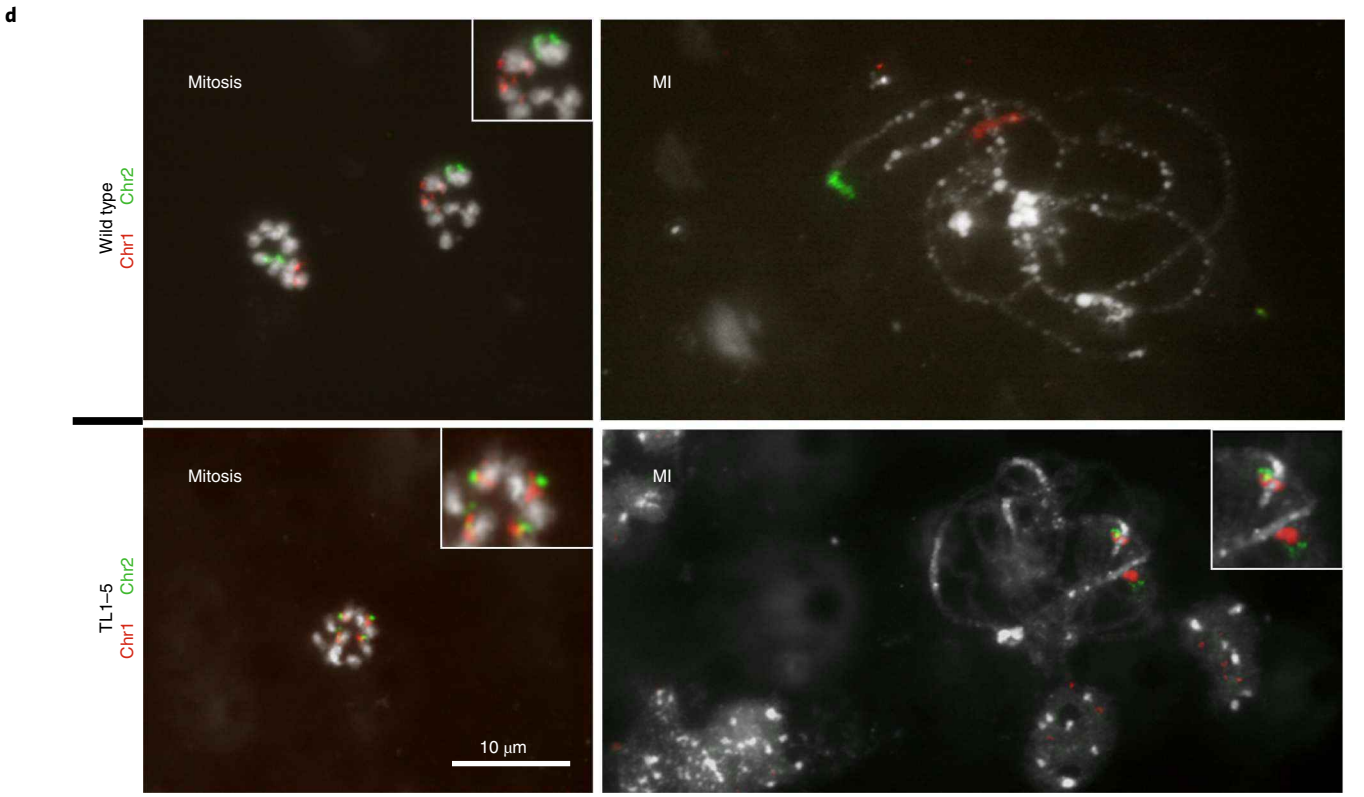


Fig. 4 | Induction of heritable translocations between Chr1 and Chr5. a, Schematic of the induction of translocations between Chr1 and Chr5 (TL1–5). Using egg-cell-specific expression of SaCas9, two DSBs were introduced on the respective chromosomes and an exchange of fragments in the Mbp range was achieved. **b**, Establishment of a line carrying the TL1–5 in the Col-0 background. Out of 40 T2 lines analysed, 1 individual plant was found that was positive for the translocation. In T3 analysis, no genotype resulting from unbalanced gametes could be detected, and 10% of the progeny carried the translocation homozygously. **c**, Sequencing analysis of the TL1–5 junctions revealed small deletions at both junction sites. **d**, Mitotic and naturally extended pachytene chromosomes of the wild type and TL1–5 after FISH using differently labelled DNA (oligonucleotide-based painting probes) specific for translocation-breakpoint regions of Chr1 (green) and Chr5 (red). Insets show further magnified chromosome regions. Experiments were repeated three times independently with similar results.

(EC1.1–EC1.2 fusion[23]) combined with the rbcS-E9 terminator (Supplementary Fig. 1). For transformation in T-DNA insertion lines of *A. thaliana*, the resistance cassette was changed from kanamycin to bar or gentamycin using PmeI and SbfI restriction sites. Spacer sequences were introduced as annealed oligonucleotides (Supplementary Table 1) into the entry vector by classical cloning with BbsI. The customized RNA chimaera is driven by the *Arabidopsis* U6-26 promoter.

For targeting two loci simultaneously, two programmed sgRNA cassettes were integrated into the destination vectors. The first chimaera was transferred using Bsu36I and MluI and the second chimaera was transferred using a Gateway LR reaction as previously described[28]. For induction of the translocation between Chr1 and Chr2 (TL1–2), protospacers (PS) 1–1 and PS2 were combined. The combination of PS1–2 and PS3 induces breaks on Chr1 and Chr5 leading to translocation TL1–5.

Plant material, growth conditions and transformation. All *Arabidopsis* lines used in this study are in the Col-0 ecotype background. In addition to the wild type, the *ku70-1* T-DNA insertion line (SALK_123114)[29] was obtained from the SALK collection[30]. For cultivation in the greenhouse, plants were grown on a substrate containing 1:1 Floraton 3 (Floragard Vertriebs GmbH) and Vermiculite (Deutsche Vermiculite Dämmstoff GmbH). For cultivation under axenic conditions, seeds were sown on agar plates containing germination medium (4.9 g l^{-1} Murashige and Skoog medium, 10 g l^{-1} saccharose, pH 5.7 and 7.6 g l^{-1} plant agar) and placed in a growth chamber. Cultivation took place at 22 °C with a 16 h:8 h light:dark cycle. For sterile plant culture, the seeds were surface sterilized with 6% sodium hypochlorite and stratified overnight at 4 °C. *Agrobacterium*-mediated transformation of plants was performed by floral dip[31].

Quantification of translocation frequencies by ddPCR. For quantification of translocation frequencies, T1 seeds were sown on plates containing germination medium, cefotaxime and gentamycin, or phosphinotricin for the selection of the transformed plants. For each line and construct, 100 whole primary transformants were selected after 14 d of growth and pooled for DNA extraction as described previously[10]. Analysis by ddPCR was performed as a probe-based duplex assay measuring the junctions in one reaction with a normalization control. Primers, dual-labelled locked nucleic acid probes and conditions for ddPCR are summarized in Supplementary Tables 2–4. The ddPCR was conducted using the QX200 AutoDG Droplet Digital PCR system, reagents, plates and cartridges from BioRad. Six biological replicates of each line were analysed. For each individual sample, the PCR mastermix was of a sufficient volume to divide into eight wells, enabling a reliable detection of rare events due to analysis of a larger amount of DNA. Subsequent analysis was completed using QuantaSoft Analysis Pro software from BioRad. A no-template control was conducted on each plate and considered in the calculation of the translocation frequency. To set thresholds for ddPCR, experiments were performed with constructs containing Gibson assembly-cloned J1 and J2 (vector backbone pEn-Chimera*BbsI, primers used listed in Supplementary Table 2).

NGS. The genomic DNA extracted from a pool of 100 primary transformants used in ddPCR was also used for analysis by NGS. To test for Cas9 efficiency, primers with 6-bp overhangs (Supplementary Table 2) were designed to amplify over the nuclease cutting sites using the Q5 High-Fidelity DNA polymerase (New England Biolabs). The 300–440-bp amplicons were purified using the peqGOLD Cycle Pure Kit (Peqlab), pooled and sent for sequencing with the Illumina HiSeq platform at GATC Biotech. For amplicon deep sequencing of the translocation junction sites, an Illumina adapter was added to the primers (Supplementary Table 2). Amplification of amplicons spanning 440–610 bp was also performed with the Q5 polymerase. The PCR reactions were sent directly to GATC Eurofins for individual sequencing on the Illumina MiSeq platform. For tag sorting and merging of paired reads, the CLC Genomics Workbench (Qiagen Bioinformatics) was used. The following data analysis was performed using the CRISPR RGEN Tool[32] with the parameters minimum frequency (n) = 1 and comparison range R = n.a. as well as wild-type marker r = n.a. for Cas9 efficiency analysis, and n = 0, R = 100 bp and r = 5 for junction-site analysis, followed by analysis using R Studio and Excel.

Line establishment for generation of homozygous translocation-carrying plants. Primary transformants (T1) were selected on agar plates containing germination medium, cefotaxime and the respective antibiotics (phosphinotricin or gentamycin) for further cultivation in the greenhouse. T2 progeny were then analysed for the translocation by PCR-based genotyping as described previously[14]. Primers used in this study are listed in Supplementary Table 2. From every selfed T1 plant, 80 seeds were sown on germination medium, and DNA was extracted from pools of 40 plantlets. These collective samples were screened for translocation events by PCR, and the lines positive for both reciprocal junctions were examined further. Individual plants positive for both junctions in the PCR screening were cultivated in the greenhouse and the resulting T3 lines were tested for segregation of the translocation events by PCR. Four different PCRs were conducted, two amplifying the wild type over the break points on both chromosomes and two amplifying the potential junction sites (Fig. 2b; primer combinations and PCR conditions in Supplementary Table 5). All plants harbouring the translocation homozygously were propagated for further analysis. To determine ligation outcomes at the junction sites on sequence level, the sites were amplified using the Q5 High-Fidelity DNA polymerase (New England Biolabs), purified with the Cycle Pure Kit (VWR International) and subsequently sent for Sanger sequencing. For detection of the translocated chromosome parts in homozygously established lines, for every 100 Mbp, a fragment of 2,000–2,500 bp was amplified by PCR (Supplementary Table 6).

Phenotypic analysis and fertility assays. To analyse phenotypic differences and fertility, plants were grown in the greenhouse for 5–6 weeks. After representative pictures were taken, five mature siliques of five plants per line were collected and incubated overnight in 70% EtOH. Determination of silique length and number of seeds per silique was performed using a binocular microscope.

FISH. FISH analysis was performed as described previously[33] using differently labelled DNA probes specific for translocation-breakpoint regions. Pools of contiguous 9 and 5 BAC clones (Supplementary Table 7) spanning 0.8 and 17.6 Mbp of Chr1 and Chr2, respectively, were used to paint the neighbouring regions of the chromosome break points during mitotic and meiotic divisions of translocation line TL1–2. Pools of 43–47-nucleotide-long oligonucleotide-based painting probes[34] (https://arborbiosci.com) spanning a region of 1 Mbp each were used to label the neighbouring regions of TL1–5 break points. An epifluorescence microscope (BX-61, Olympus) equipped with UPlan(F), ×100, 1.30 numerical aperture objective lens (Olympus) and a cooled black and white CCD camera (ORCA-R2, Hamamatsu) under the control of the microscope software CellSens Dimension (Olympus) were used to record the micrographs. The following fluorescence filters were used: FITC-2024B-000, Txred-4040C-000n and Sp. Aqua HC (Sembrock). Black and white pictures were pseudocoloured with the software Adobe Photoshop.

Statistical analysis. Statistical analyses were performed using R or Excel. To determine statistical significance levels of translocation frequencies, a two-tailed Mann–Whitney U-test with unequal variance was performed. For comparison of fertility, an unpaired two-sample t-test with unequal variance was performed. $^{***}P < 0.001$, extremely significant; $0.001 \leq {}^{**}P < 0.01$, very significant; $0.01 < {}^{*}P \leq 0.05$, significant; $P > 0.05$, not significant.

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 the data supporting the findings are available within the paper and its Supplementary Information, or are available from the corresponding author upon reasonable request.

Code availability

R code for detailed analysis of NGS data is available upon request from the corresponding author.

Received: 4 February 2020; Accepted: 14 April 2020;
Published online: 25 May 2020

References

1. Zhang, Y., Massel, K., Godwin, I. D. & Gao, C. Applications and potential of genome editing in crop improvement. *Genome Biol.* **19**, 210 (2018).
2. Zhang, Y., Malzahn, A. A., Sretenovic, S. & Qi, Y. The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants* **5**, 778–794 (2019).
3. Schindele, A., Dorn, A. & Puchta, H. CRISPR/Cas brings plant biology and breeding into the fast lane. *Curr. Opin. Biotechnol.* **61**, 7–14 (2020).
4. Hua, K. et al. Perspectives on the application of genome-editing technologies in crop breeding. *Mol. Plant* **12**, 1047–1059 (2019).
5. Steinert, J., Schiml, S., Fauser, F. & Puchta, H. Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*. *Plant J.* **84**, 1295–1305 (2015).
6. Schmidt, C., Schindele, P. & Puchta, H. From gene editing to genome engineering: restructuring plant chromosomes via CRISPR/Cas. *aBIOTECH* **36**, 17 (2019).
7. Wolter, F., Schindele, P. & Puchta, H. Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol.* **19**, 176 (2019).
8. Puchta, H. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* **56**, 1–14 (2005).
9. Filler Hayut, S., Melamed Bessudo, C. & Levy, A. A. Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nat. Commun.* **8**, 15605 (2017).
10. Salomon, S. & Puchta, H. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J.* **17**, 6086–6095 (1998).
11. Schimmel, J., Kool, H., van Schendel, R. & Tijsterman, M. Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *EMBO J.* **36**, 3634–3649 (2017).

12. Wang, H. & Xu, X. Microhomology-mediated end joining: new players join the team. *Cell Biosci.* **7**, 6 (2017).
13. Siebert, R. & Puchta, H. Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell* **14**, 1121–1131 (2002).
14. Schmidt, C., Pacher, M. & Puchta, H. Efficient induction of heritable inversions in plant genomes using the CRISPR/Cas system. *Plant J.* **98**, 577–589 (2019).
15. Pacher, M., Schmidt-Puchta, W. & Puchta, H. Two unlinked double-strand breaks can induce reciprocal exchanges in plant genomes via homologous recombination and nonhomologous end joining. *Genetics* **175**, 21–29 (2007).
16. Rowley, J. D. Chromosome translocations: dangerous liaisons revisited. *Nat. Rev. Cancer* **1**, 245–250 (2001).
17. Thompson, S. L. & Compton, D. A. Chromosomes and cancer cells. *Chromosome Res.* **19**, 433–444 (2011).
18. Bunting, S. F. & Nussenzweig, A. End-joining, translocations and cancer. *Nat. Rev. Cancer* **13**, 443–454 (2013).
19. Lysak, M. A. et al. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl Acad. Sci. USA* **103**, 5224–5229 (2006).
20. Gabur, I., Chawla, H. S., Snowdon, R. J. & Parkin, I. A. P. Connecting genome structural variation with complex traits in crop plants. *Theor. Appl. Genet.* **132**, 733–750 (2019).
21. Huefner, N. D., Mizuno, Y., Weil, C. F., Korf, I. & Britt, A. B. Breadth by depth: expanding our understanding of the repair of transposon-induced DNA double strand breaks via deep-sequencing. *DNA Repair* **10**, 1023–1033 (2011).
22. Graham, T. G. W., Carney, S. M., Walter, J. C. & Loparo, J. J. A single XLF dimer bridges DNA ends during nonhomologous end joining. *Nat. Struct. Mol. Biol.* **25**, 877–884 (2018).
23. Wang, Z.-P. et al. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome Biol.* **16**, 144 (2015).
24. Wolter, F., Klemm, J. & Puchta, H. Efficient in planta gene targeting in *Arabidopsis* using egg cell-specific expression of the Cas9 nuclease of *Staphylococcus aureus*. *Plant J.* **94**, 735–746 (2018).
25. Durr, J., Papareddy, R., Nakajima, K. & Gutierrez-Marcos, J. Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in *Arabidopsis* using CRISPR/Cas9. *Sci. Rep.* **8**, 4443 (2018).
26. Miki, D., Zhang, W., Zeng, W., Feng, Z. & Zhu, J.-K. CRISPR/Cas9-mediated gene targeting in *Arabidopsis* using sequential transformation. *Nat. Commun.* **9**, 1967 (2018).
27. Kazda, A. et al. Chromosome end protection by blunt-ended telomeres. *Genes Dev.* **26**, 1703–1713 (2012).
28. Schiml, S., Fauser, F. & Puchta, H. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant J.* **80**, 1139–1150 (2014).
29. Jia, Q., Bundock, P., Hooykaas, P. J. J. & Pater, Sde *Agrobacterium tumefaciens* T-DNA integration and gene targeting in *Arabidopsis thaliana* non-homologous end-joining mutants. *J. Bot.* **2012**, 1–13 (2012).
30. Alonso, J. M. et al. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).
31. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
32. Park, J., Lim, K., Kim, J.-S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33**, 286–288 (2017).
33. Lysak, M. A., Fransz, P. F., Ali, H. B. & Schubert, I. Chromosome painting in *Arabidopsis thaliana*. *Plant J.* **28**, 689–697 (2001).
34. Han, Y., Zhang, T., Thammapichai, P., Weng, Y. & Jiang, J. Chromosome-specific painting in *Cucumis* species using bulked oligonucleotides. *Genetics* **200**, 771–779 (2015).

Acknowledgements

We thank A. Imhof, K. Siegmund, K. Kumke and C. Gäde for technical assistance with performing experiments, T. Wehrle for assisting with cultivation of the plants in the greenhouse, A. Whitbread for critical reading of the manuscript, T. Zundel for providing an R-based script for bioinformatic analysis of NGS data, A. Neubauer for customizing the script to our needs and M. Lysak for the selection of BACs. This research was funded by the European Research Council (ERC) (Grant number ERC-2016-AdG_741306 CRISBREED) to H.P.

Author contributions

N.B., C.S., M.P. and H.P. designed research; N.B. conducted the research and A.H. performed FISH analysis; N.B., C.S., A.H. and H.P. analysed data; and N.B., A.H. and H.P. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-020-0663-x.

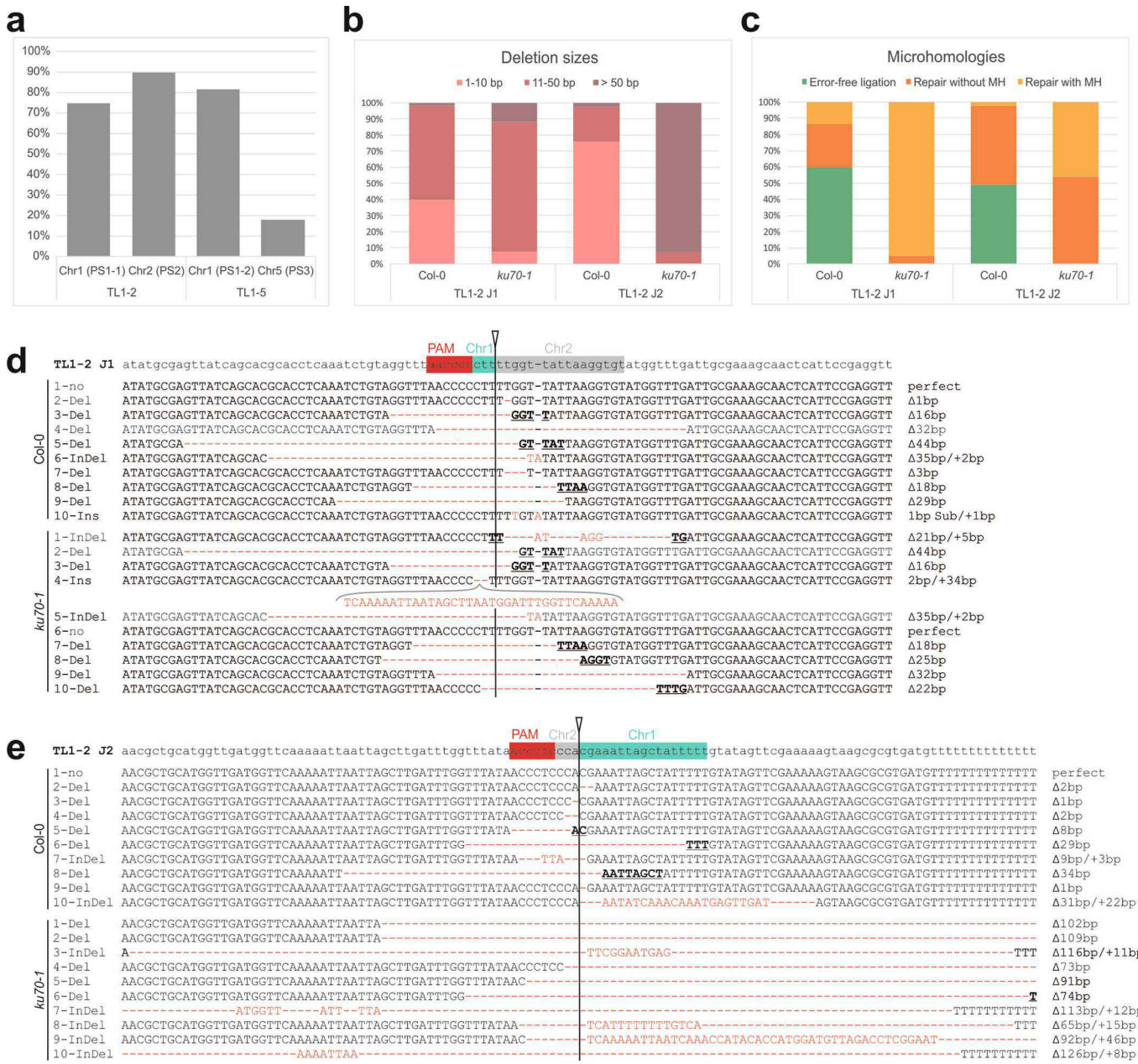
Supplementary information is available for this paper at https://doi.org/10.1038/s41477-020-0663-x.

Correspondence and requests for materials should be addressed to H.P.

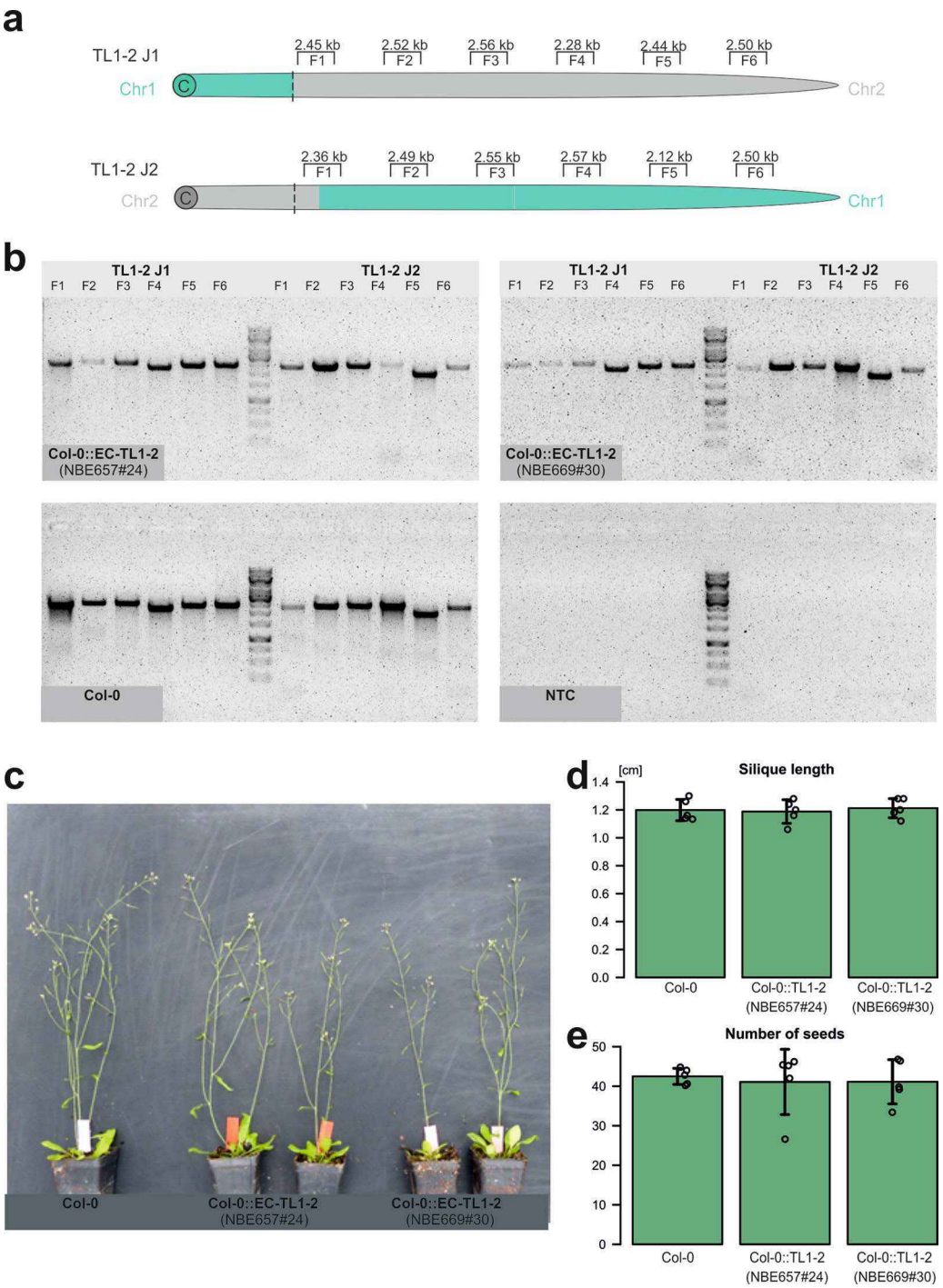
Peer review information *Nature Plants* thanks Keunsub Lee, Kan Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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 | Detailed analysis of NGS data regarding the repair pattern at the junction sites of TL1-2 in T1 generation. a, Editing efficiency (percentage of modified reads out of total amount of reads) of SaCas9 at the four different target sites. **b**, Evaluation of NGS data regarding the deletions at the junction sites. Of this, only the reads carrying a deletion were analysed. The lengths of the deletions were divided into three classes (small: 1-10 bp, middle: 11-50 bp, large: > 50 bp). The deletions at the junction sites in Col-0 were mostly of small or middle size, whereas the *ku70* mutant showed bigger amounts of large deletions at both junction sites. **c**, Quantification of the occurrence of microhomologies (MH, ≥ 2 bp) used for junction formation. We separated the reads into three classes: error-free ligated junctions (green), junctions formed without the use of MHs (orange) or with MHs (yellow). In wild type, most junctions were directly ligated without any mutation induction at the junction sites. Of the reads showing mutations at the junction sites, most junctions were joined without the use of MHs, only a minority of junctions were joined using MHs. In contrast, in the *ku70* mutant background nearly no error-free ligation occurred and the prevalent repair pattern at the junctions sites showed the use of MHs for joining. **d, e**, Detailed representation of the 10 most common reads of both junction sites on sequence level in wild type as well as *ku70-1*.



Extended Data Fig. 2 | Molecular and cytological analysis of two independent plant lines (NBE657#24 and NBE669#30) carrying the TL1-2 homozygously. a, Schematic overview of PCR amplicons (F1-6) of the translocated chromosome parts. A fragment for amplification of around 2.5 kb every 100 kb was designed for each chromosome arm. **b**, Cropped gel electrophoresis pictures of PCR amplicons of the translocated chromosome parts. Every 100 kb on the translocated chromosome parts, a band could be detected, indicating no information was lost during translocation formation. As PCR controls, genomic DNA of wild type without translocation formation (positive control) and water (NTC = no template control, negative control) were processed at the same time and loaded on different gels for a better overview. **c**, By phenotypical comparison of 5 week old plants carrying the TL1-2 to the wild type no differences in growth could be documented. Experiments were repeated two times independently with similar results. **d, e**, Fertility analysis was conducted by measuring the silique length and counting the number of seeds of five biologically independent samples ($n = 5$). Barplots show the mean values, error bars as mean ± s.d. For both analysed lines carrying the TL1-2, we could not detect any difference to the wild type.

nature research

Corresponding author(s): Holger Puchta

Last updated by author(s): Mar 23, 2020

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 Authors & Referees 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 custom code was used in data collection. All sequence information was obtained from TAIR (https://www.arabidopsis.org/). Sanger and next generation sequencing was performed by GATC Eurofins. The automated imaging system used for gel documentation was GEL iX Imager (Intas). As an epifluorescence microscope the BX-61 (Olympus) equipped with UPlan(F), 100x/1.30 objective lens (Olympus) and a cooled black/white CCD camera (ORCA-R2, Hamamatsu) under the control of the microscope software CellSens Dimension 1.11 (Olympus) was used with the following fluorescence filters: FITC-2024B-000, Txred-4040C-000n and Sp. Aqua HC (Sembrock). Please see material and methods for details.
Data analysis	During the analysis we used published or open source software as well as commercial software. For evaluation of droplet digital PCR results QuantaSoft Analysis Pro 1.0 (BioRad) was used. Raw reads were processed through CLC Genomics workbench (version 10.1.11). Subsequent analysis was done with the Cas9 analyzer (http://www.rgenome.net/cas-analyzer/#!) followed by Rstudio (version 1.1.463) with our own custom code which is available upon request to the corresponding author. Graphs were made in Rstudio (1.1.463) for boxplot and Excel 2016 for bar charts. ApE (v2.0.55) was used for alignment and analysis of Sanger sequencing data. Brightness and contrast of gel electrophoresis pictures were adjusted in Adobe Photoshop CC (14.1.0.401). Black and white microscopy pictures were pseudo-coloured with the software Adobe Photoshop 6 (21.0.2).

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/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

The authors declare that the data supporting the findings are available within the paper and its Supplementary Information, or are available from the corresponding author upon reasonable request. Deep sequencing data that support the findings of this study have been deposited in SRA [accession: PRJNA607494]. Raw data for Figure 1 and Supplementary Figure 1 are listed in Supplementary Data 1-4.

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	No sample-size calculation was performed. As translocation frequencies were expected to be very low and to randomize bias from integration of the Cas9 nuclease, rather large sample sizes were chosen. For T1 analysis, six biological replicates with 100 T1 plants each were analysed for both lines. According to our previous experiece in the Arabidopsis system and as no discordant values were detected, this sample size is sufficient to ensure reproducibility and detect significant differences.
Data exclusions	No data has been excluded.
Replication	All attempts at replication were successful. True biological replicates (i.e., independent plants) were used as replicates for statistical analyses. The number of replicates is given in the Figure legends. For deep sequencing, the experiment was performed only once with one biological replicate out of genotyping costs. To ensure reproducibility, the plants were grown in a growth chamber under well defined conditions and samples were taken at the same developmental stage.
Randomization	All independent transgenic lines were randomly picked, and all individuals for analysis were likewise randomly picked.
Blinding	Not required for most analyses, as samples were processed identically through standard and in some cases automated procedures (ddPCR, DNA sequencing) that should not bias outcomes. Phenotypical analysis and fertility assays were done blinded by randomly numbering the plant lines and assigning the genotype after analysis.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*

Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used — *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation — *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) — *State the source of each cell line used.*

Authentication — *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination — *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See ICLAC register) — *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology

Specimen provenance — *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*

Specimen deposition — *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods — *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals — *For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*

Wild animals — *Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Field-collected samples — *For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Human research participants

Policy information about studies involving human research participants

Population characteristics — *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic*

Population characteristics	*information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

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	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

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

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.