

An active Helitron transposon family in wheat

Received: 30 June 2025

Accepted: 12 May 2026

Published online: 5 June 2026

 Check for updates

Haoran Peng (彭浩然)^{1,2,8}, Lijing Tang (唐粒菁)^{1,3,8}, Nataliya Hrunyk¹, Roman Kellenberger¹, Dario Fossati⁴, Hélène Rimbart⁵, Pierre Sourdille⁵, Shengwei Ma (马省伟)⁶, Alison B. Hickman⁷, Fred Dydá⁷, Claudia Köhler², Carole Charlier³, Frédéric Choulet⁵ & Etienne Bucher¹✉

Transposable elements play a pivotal role in genome evolution and phenotypic variation in numerous eukaryotic species¹. Helitrons, a recently identified category of transposons, remain poorly understood in terms of epigenetic regulation and real-time mobilization in plants^{2,3}. Here our study reveals that reduced DNA methylation combined with heat stress promotes the mobilization of the *Xuan–Feng* Helitron family in wheat. Activation is marked by transcription, extrachromosomal circular DNA formation and novel somatic insertions. Genetic segregation and heterologous reconstitution establish *Feng8* as the autonomous driver of the *Xuan–Feng* family mobilization. These findings represent a step forward in the study of active Helitrons and their potential biological functions as well as their role in genome dynamics and their potential use in crop breeding.

The genomes of most plant species predominantly consist of transposable elements (TEs)⁴. Hexaploid bread wheat (*Triticum aestivum*), the most important crop worldwide, exhibits a large genome size (16 Gb), with more than 85% originating from TEs and other repetitive sequences⁵. The significance of TEs in the functional and evolutionary dynamics of genomes is now widely acknowledged. TEs contribute to the regulation of gene expression, stress responses and adaptation, and they played key roles during crop domestication^{6–9}. Helitrons, a distinct category of DNA TEs, eluded discovery until the early 2000s when bioinformatic methodologies enabled their identification over 50 years after the initial discovery of TEs^{2,10}. Helitrons have been suggested to employ a rolling-circle transposition mode reminiscent of plasmid replication and single-stranded DNA viruses¹¹. This transposition mode is distinct, as it does not generate target site duplications, a characteristic footprint of many TEs¹².

A canonical Helitron has conserved termini: a 5' TC, a 3' CTRR sequence (where R stands for A or G) and a 16–20-bp hairpin structure located 10–15 bp from the 3' end^{2,13}. Due to the high diversity of their

internal sequences, Helitron families are defined according to sequences of 30 bases that are present in the termini and referred to as the left terminal sequence (LTS) and the right terminal sequence (RTS)^{14,15}. Whereas their relics are abundant across the eukaryotic kingdom and indicative of historical activity¹⁶, direct evidence of real-time Helitron mobility is scarce¹⁷. Indeed, controllable active TE mobility is rarely observed in plants, largely due to the concerted action of multiple epigenetic mechanisms that silence them^{18,19}; for example, DNA methylation and small RNA pathways robustly suppress their mobility^{1,20}. Given that Helitrons employ extrachromosomal circular DNA (eccDNA) intermediates during their transposition process^{21,22}, these eccDNAs may serve as valuable diagnostic markers for assessing Helitron activity. To promote TE mobilization using a similar approach as previously reported²³, we treated wheat seedlings with a combination of two chemicals that perturb transcriptional and epigenetic homeostasis: the RNA polymerase I, II and III inhibitor juglone (J)^{24,25} and the DNA methylation inhibitor zebularine (Z)^{26,27}. Since numerous

¹Crop Genome Dynamics Group, Agroscope, Nyon, Switzerland. ²Department of Plant Reproductive Biology and Epigenetics, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany. ³Unit of Animal Genomics, GIGA & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium.

⁴Field Crop Breeding and Genetic Resources, Agroscope, Nyon, Switzerland. ⁵INRAE, GDEC, Université Clermont Auvergne, Clermont-Ferrand, France.

⁶Yazhouwan National Laboratory, Sanya, China. ⁷Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. ⁸These authors contributed equally: Haoran Peng, Lijing Tang. ✉e-mail: etienne.bucher@agroscope.admin.ch

TEs are stress-responsive²⁸, we combined the drug treatments (JZ) with heat stress (H). In the wheat cultivar Arina, the JZH-treated seedlings exhibited substantially reduced vigour, characterized by shorter roots and delayed leaf development (Fig. 1a). Nanopore-based whole-genome sequencing and DNA methylation calling revealed that global DNA methylation across the three sequence contexts (CG, CHG and CHH) was reduced following JZH treatment. In the CG, CHG and CHH methylation contexts, the levels decreased by 15%, 9.3% and 14%, respectively, compared with control plants (Extended Data Fig. 1a). To identify activated Helitrons, eccDNAs from these plants were extracted, amplified by rolling circle amplification (RCA) and subjected to long-read sequencing^{29,30}. We observed that JZH-treated samples generated a unique eccDNA, precisely overlapping with a Helitron TE annotation (2,513 bp, chromosome 6D: 152366979–152369491 of the ArinaLrFor reference genome³¹), whereas this signal was absent in the control samples (Extended Data Fig. 2a). Dot-plot graphs generated from the long reads revealed numerous reads containing multiple tandem repeat copies of the Helitron sequence. These repeats were the result of the in vitro RCA reaction that was used to amplify the eccDNA signal (Extended Data Fig. 2b), thus confirming the circular nature of the original DNA templates.

Next, to independently confirm the presence of Helitron eccDNA, we conducted inverse PCR using primers designed to span the eccDNA junction (Fig. 1b). We ruled out any non-specific signals potentially originating from genomic tandem repeat sequences in the Arina genome, as we did not detect any PCR products when genomic DNA was used as a PCR template (Extended Data Fig. 2c). To determine the requirements for efficient Helitron eccDNA production, we subjected plants to individual treatments with drugs alone (JZ), heat stress alone (H) and their combination (JZH). Helitron-derived eccDNA was exclusively detected via inverse PCR on RCA-amplified DNA from the JZH group (Extended Data Fig. 2d). These results demonstrate that Helitron-derived eccDNA accumulation is heat stress responsive but repressed by DNA methylation and silencing.

We then wanted to test whether eccDNA accumulation varied between distantly related wheat cultivars. We tested four European cultivars (Arina, Renan, CH Nara and Rosatch) and two Asian cultivars (Chinese Spring (CS) and Norin61). JZH-treated Arina, Renan, CH Nara and Rosatch produced detectable Helitron-derived eccDNAs, but the untreated control groups did not (Fig. 1b). Conversely, the cultivars CS and Norin61 did not produce detectable levels of Helitron-derived eccDNAs following JZH treatment (Fig. 1b). This demonstrates that not all wheat varieties have the capacity to activate this element even though it is present in their genome. To delineate the exact junction between the RTS and LTS of this Helitron in the varieties that showed eccDNA accumulation, we subjected the obtained PCR fragments to Sanger sequencing. We found that the junctions presented the exact expected RTS and LTS junctions in these varieties (Extended Data Fig. 2e).

The eccDNA-producing Helitron exhibited canonical features, initiating with a TC dinucleotide and terminating with a CTAG tetranucleotide (Fig. 1c). We also identified a 17-nucleotide hairpin structure, including a 5-nucleotide loop and a 6-nucleotide stem, located 11 nucleotides upstream of the RTS (Extended Data Fig. 3a). On the basis of the absence of transposase coding capability (Extended Data Fig. 3a), we concluded that this element was a non-autonomous Helitron and designated it *Xuan1*. Non-autonomous TEs do not encode a functional transposase necessary for self-transposition. Instead, they rely on *trans*-mobilization facilitated by the enzymatic machinery of their autonomous counterparts from the same family¹². To delineate this Helitron family, we employed a classification approach based on the 30-bp LTS and RTS regions^{14,15}. We designated this family *Xuan–Feng* (meaning ‘whirlwind’ in Chinese; *Xuan* signifies non-autonomous family members, and *Feng* denotes potentially autonomous ones). The annotation of this family identified *Xuan1–77* and *Feng1–7* (Supplementary Table 1). In addition, by employing whole-genome

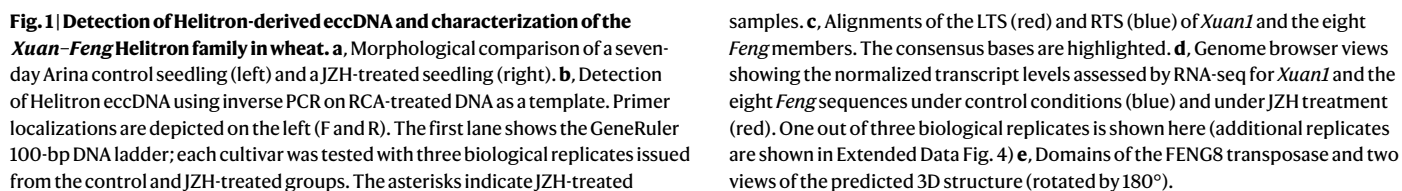
sequencing with Nanopore reads, we discovered *Feng8*, which was not completely assembled in the ArinaLrFor reference genome³¹ (Fig. 1c, Extended Data Fig. 3b and Supplementary Table 1).

The *Xuan1* element and putatively autonomous *Feng* members exhibited conserved LTS and RTS, except for *Feng6* and *Feng7*, which had accumulated mutations in their LTS (Fig. 1c). Using a multiple alignment approach, *Feng* members were clustered into two groups: one group that included *Feng1*, *Feng2* and *Feng7* and a second group that contained all other *Feng* copies. Notably, *Feng4*, *Feng6* and *Feng8* showed over 99% nucleotide sequence identity (Extended Data Fig. 3c). We also detected a predicted heat response element with the consensus sequence nTTCnnGAAn in the promoter regions of *Feng3–6* and *Feng8* (Extended Data Fig. 3d). Notably, the *Arabidopsis* retrotransposon *ONSEN* (*ATCOPIA78*) was found to include the same sequence motif in its heat-stress-responsive promoter³². We also observed a preference for insertions of this Helitron family at AT dinucleotides (Extended Data Fig. 3e), which is a characteristic of Helitrons observed across different species¹¹.

To identify the autonomous *Feng* copy that facilitated the biosynthesis of *Xuan1* eccDNA, we examined the transcriptional activity of all *Feng* members using RNA sequencing (RNA-seq) to determine whether they responded to the JZH treatment. We observed only minimal, statistically non-significant, transcript levels for *Xuan1* in response to the JZH treatment (Fig. 1d). By contrast, the *Feng* elements displayed varying transcriptional responses following JZH treatment. *Feng2* and *Feng5* did not exhibit significant transcriptional activation, while *Feng1*, *Feng3*, *Feng4* and *Feng6* to *Feng8* showed significantly increased transcript levels upon JZH treatment (Fig. 1d and Extended Data Fig. 4). *Feng1* produced only a short 1,284-bp transcript following JZH treatment. Although *Feng3* and *Feng7* exhibited a partial constitutive transcriptional signal under control conditions, complete transcripts and significantly increased transcript levels were only observed in the JZH-treated group (Fig. 1d and Extended Data Fig. 4). A close inspection of the transcripts produced by *Feng3*, *Feng4*, *Feng6*, *Feng7* and *Feng8* revealed that all *Feng* copies, except *Feng8*, carried nonsense mutations or a frameshift mutation in the coding sequence (Extended Data Fig. 5a). We thus concluded that only *Feng8* had the capacity to code for a functional transposase. Next, we examined whether DNA methylation levels are affected at *Feng8* in JZH-treated plants. In untreated plants, the region was predominantly methylated in the CG context. Following JZH treatment, a subset of CG and CHG sites showed reduced methylation levels, consistent with increased transcriptional activity (Fig. 1d and Extended Data Fig. 1b). In contrast, the *Xuan1* locus exhibited low levels of DNA methylation across all sequence contexts in untreated plants, and this hypomethylated state was maintained after JZH treatment (Extended Data Fig. 1c).

On the basis of the RNA-seq data, we found that *Feng8* contained a coding sequence spanning 5,028 bp and encoded a transposase of 1,676 amino acids (Fig. 1e and Extended Data Fig. 5). This coding sequence consists of seven exons, and the first five introns start with a GT donor site and end with the AG acceptor site, whereas the last intron starts with GC. To verify the functional protein domains, we aligned FENG8 with the Helraiser protein (an in silico reconstituted Helitron transposase from *Myotis lucifugus*)³³. We noticed that the FENG8 transposase had a larger N-terminal domain (ND) (Met1–Leu638) and harboured a potential zinc-finger-like motif (Arg311–Val337). The RepHel core of FENG8 comprised a HUH endonuclease domain spanning 289 amino acids (Pro668–Met956), featuring the conserved HUH motif (His767 and His769), as well as two tyrosine residues (Tyr902 and Tyr906) known to form the active site in Helraiser. Additionally, it encompassed a helicase domain spanning 483 (Arg1194–Val1676) amino acids, containing the eight conserved motifs that are characteristic of the SF1 superfamily of DNA helicases (Extended Data Fig. 5b).

The structure predicted by AlphaFold2 (ref. 34) revealed the presence of an ND and a helicase domain connected by two intermediate



domains, INT1 and INT2, as observed in the experimentally determined structure of Helraiser, and similarly forming a large, tightly knit molecule (Fig. 1e and Extended Data Fig. 6a). The first 264 amino acids of the ND are predicted to be disordered except for two α -helices (Extended Data Fig. 6b). This is consistent with the cryo-electron microscopy structure of Helraiser, where the first 110 amino acids were found to be disordered³³. Given the amino acid identity (33.92%) between FENG8 and Helraiser, the predicted and experimental protein structures both exhibited high overall similarity, suggesting a similar function (Extended Data Fig. 6c–e).

To investigate the prevalence of the FENG8 transposase in other species, we extracted a total of 83 Helitron-like element coding sequences from the NCBI database plus Helraiser (Supplementary Table 2). The phylogenetic tree indicated that FENG8 is more closely related to Helitrons present in monocotyledonous plants, whereas Helraiser clustered with animal and insect sequences, suggesting a long-standing evolutionary divergence (Extended Data Fig. 7).

We next wanted to genetically test whether *Xuan1* mobilization requires *Feng8*. To achieve this, we took advantage of the contrasting responses of two wheat varieties to the JZH treatments: first, the variety CS, in which *Feng8* is absent³⁵ and which did not spawn *Xuan1* eccDNA following JZH treatment (Fig. 1b); and second, the Renan cultivar, which does carry an intact *Feng8* copy in its genome³⁶ and which spawned *Xuan1* eccDNAs upon JZH treatment (Fig. 1b and Extended Data Fig. 8). We cloned and sequenced *Feng8* from the Arina and Renan genomic DNA and found that they showed 100% identity in the nucleotide sequence. Similarly, our comparison of the *Xuan1* sequences of the Arina, CS and Renan varieties showed 100% identity. To genetically test the requirement of *Feng8* for *Xuan1* mobilization, we took advantage of wheat recombinant inbred lines (RILs) that were obtained from an initial cross of CS and Renan followed by eight generations of self-pollination (Fig. 2a). First, on the basis of single-nucleotide polymorphism (SNP) genotyping data, 21 RILs were selected: 10 lines with *Feng8* (the Renan allele in the region) and 11 lines without (the CS allele in the region). The presence or absence of *Feng8* was further confirmed by PCR for all RILs (Fig. 2b). To test the capacity of these RILs to produce *Xuan1* eccDNA, they were all subjected to JZH treatment. We found that only the RILs with *Feng8* produced *Xuan1* eccDNA, as confirmed by inverse PCR (Fig. 2c). We therefore demonstrate a genetic association between *Feng8* and the induction of *Xuan1* eccDNA biosynthesis. As eccDNA levels were very low in these lines, we consistently performed these assays with three biological replicates, and we never detected false positives.

Next, we wanted to test whether *Xuan1* and *Feng8* can complete their life cycle by integrating de novo into the wheat genome. To capture somatic de novo insertions of *Xuan1* and *Feng8* in the RILs, we harnessed the high sensitivity of the well-established pooled CRISPR inverse PCR sequencing method (PCIP-seq) (Methods and Extended Data Fig. 9a). This method has been demonstrated to efficiently detect rare somatic genomic insertions of exogenous and endogenous viruses (including TEs) in humans, cattle and sheep^{37,38}. We blindly applied this method to the 21 RILs to detect novel somatic *Xuan1* insertions. We captured a total of 49 somatic *Xuan1* de novo insertions in 8 out of 10 RILs that carried *Feng8* and none in the 11 RILs

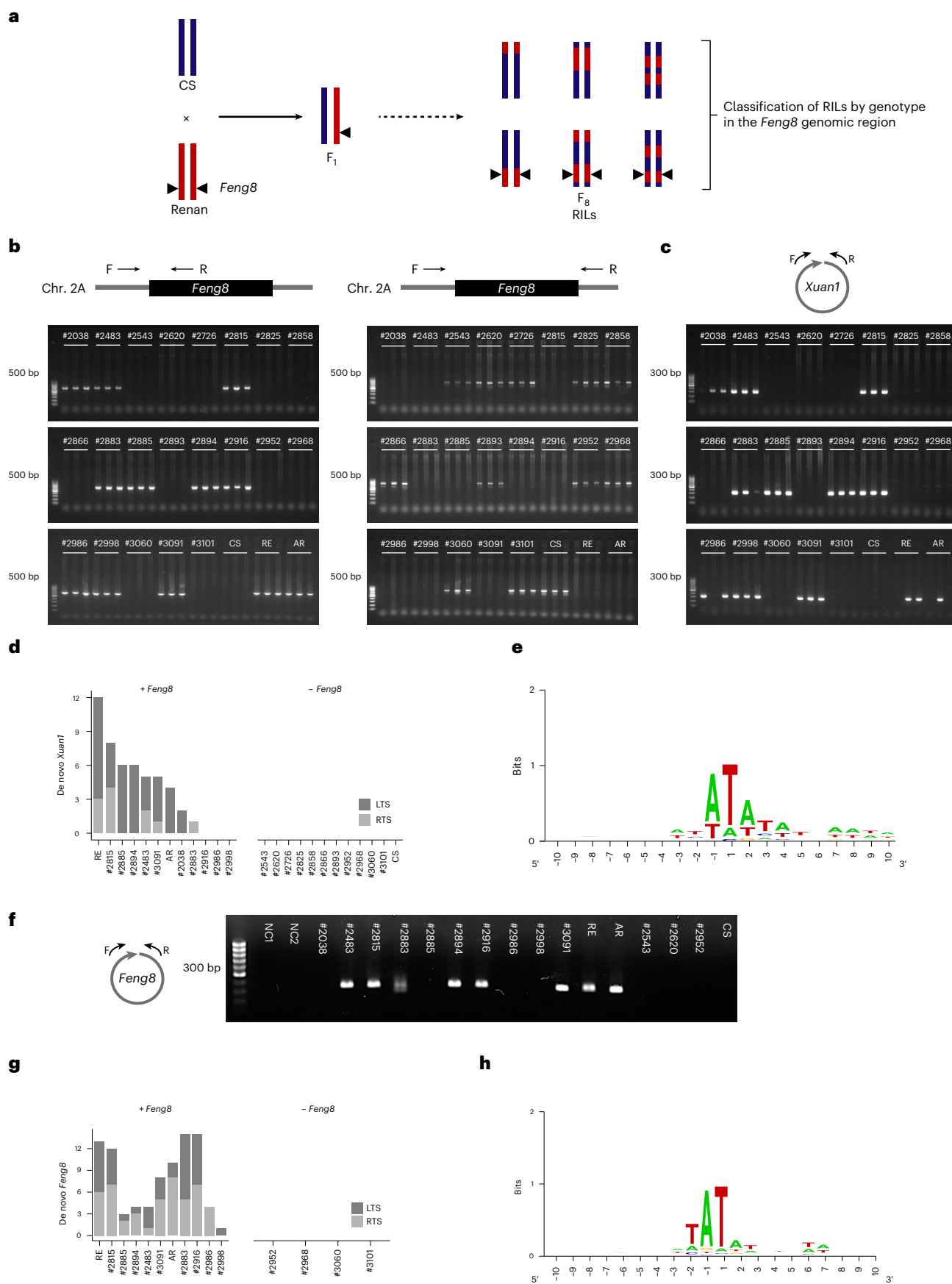
without *Feng8* (Fig. 2d and Supplementary Table 3). We extracted 20 bp of flanking wheat host DNA sequence at these 49 de novo insertion sites and found a marked preference for insertions at AT dinucleotides (Fig. 2e and Supplementary Table 3), an expected feature for bona fide Helitron insertions. To test whether *Feng8* itself was also circularized and mobilized, we applied inverse PCR to detect *Feng8* eccDNA and PCIP-seq to detect somatic *Feng8* de novo insertions in the RILs and the parental lines. We found that *Feng8* also produced eccDNA (Fig. 2f) and was mobile, as we detected 86 de novo insertions in the *Feng8*-carrying lines (Fig. 2g) and none in the lines that do not have *Feng8*. Again, we found a marked insertion preference at AT dinucleotides, demonstrating the specificity of the method (Fig. 2h). For all *Xuan1* and almost all *Feng8* insertions, we could detect either the LTS or RTS junction but not both for the same insertion event. That could be expected as we sampled the plants shortly after stress, thus following few cell divisions. Still, for one *Feng8* insertion, we managed to capture both junctions (Extended Data Fig. 9b and Supplementary Table 3). From that we conclude that *Xuan1* and *Feng8* insertions are rare and predominantly existed in only a single cell, thus providing only a single DNA fragment available for capture. Notably, we found that both *Xuan1* and *Feng8* integrated in the vicinity of or within genes in 34.6% of the cases (Supplementary Table 3). To assess the heritability of *Xuan1* or *Feng8* insertions to the next generation, we tested 99 plants descending from 6 JZH-treated mother plants for novel insertions. None of the progeny plants had inherited novel *Xuan1* or *Feng8* insertions, suggesting that the transmission of newly induced insertions to the next generation did not occur, or occurs only rarely under the conditions tested.

To further characterize the heat responsiveness of *Feng8*, 966 bp of its promoter region was fused to the β -glucuronidase (GUS) reporter gene and transformed into *Arabidopsis thaliana*. Under control conditions (21 °C), GUS activity was detected primarily in vascular tissues, the shoot apical region and embryonic tissues. Upon heat treatment at 37 °C, GUS staining intensity increased markedly and became more broadly distributed across tissues (Fig. 3a). These results indicate that the *Feng8* promoter is temperature-responsive and sufficient to drive transcriptional activation under heat stress.

To directly test whether *Feng8* is sufficient to drive *Xuan1* mobilization, we generated independent transgenic *Arabidopsis* lines containing (1) *Xuan1* alone (#X1), (2) *Feng8* alone (#F8) and (3) *Xuan1* together with *Feng8* (#X1F8). On the basis of the promoter assay results, all lines were subjected to heat treatment at 37 °C to induce *Feng8* expression. We first examined eccDNA formation by inverse PCR. #X1 lines did not produce detectable *Xuan1*-derived eccDNA. In contrast, #F8 lines showed robust *Feng8*-derived eccDNA formation, and #X1F8 lines produced both *Xuan1* and *Feng8* eccDNA (Fig. 3b). To determine whether eccDNA formation corresponds to bona fide somatic transposition events and to employ a complementary sequencing approach to PCIP, we used the transposable element display sequencing (TED-seq) protocol to detect somatic insertions³⁹. We identified one *Xuan1* insertion and three *Feng8* insertions supported by high-confidence paired-end reads, all occurring at AT dinucleotides. In addition, we manually inspected the full R1 reads and confirmed the presence of the expected Helitron terminal sequence at the insertion junctions (Fig. 3c and Supplementary Table 4). A de novo *Xuan1* insertion was detected

Fig. 2 | *Feng8* is necessary for *Xuan1* eccDNA biosynthesis and de novo insertion. **a**, The crossing scheme used to obtain CS × Renan RILs. The blue segments highlight chromosome segments originating from CS and the red ones those originating from Renan. The black triangles indicate the presence of *Feng8*. Each RIL was first analysed by SNP genotyping, and then the presence or absence of *Feng8* was confirmed by PCR (see c). **b**, *Feng8* genotyping by PCR for the presence (left) and absence (right) of the Helitron. The first lane shows the GeneRuler 100-bp DNA ladder. RE, Renan; AR, Arina. **c**, Inverse PCR detection of *Xuan1* eccDNA for each JZH-treated RIL (indicated by the numbers at the top) with three biological replicates. The first lane shows the GeneRuler 100-bp

DNA ladder. **d**, Quantification of somatic *Xuan1* de novo insertion events in the genomes of RILs that have *Feng8* (+) and lines that do not (–). **e**, Sequence logo of 49 de novo *Xuan1* integration sites (the insertion site is between the –1 and 1 base positions). **f**, Inverse PCR detection of *Feng8* eccDNA for each JZH-treated RIL (indicated by the numbers at the top). The first lane shows the GeneRuler 100-bp DNA ladder. NC, non-DNA control. **g**, Quantification of somatic *Feng8* de novo insertion events in the genomes of RILs that have *Feng8* (+) and lines that do not (–). **h**, Sequence logo of 86 de novo *Feng8* integration sites (the insertion site is between the –1 and 1 base positions).



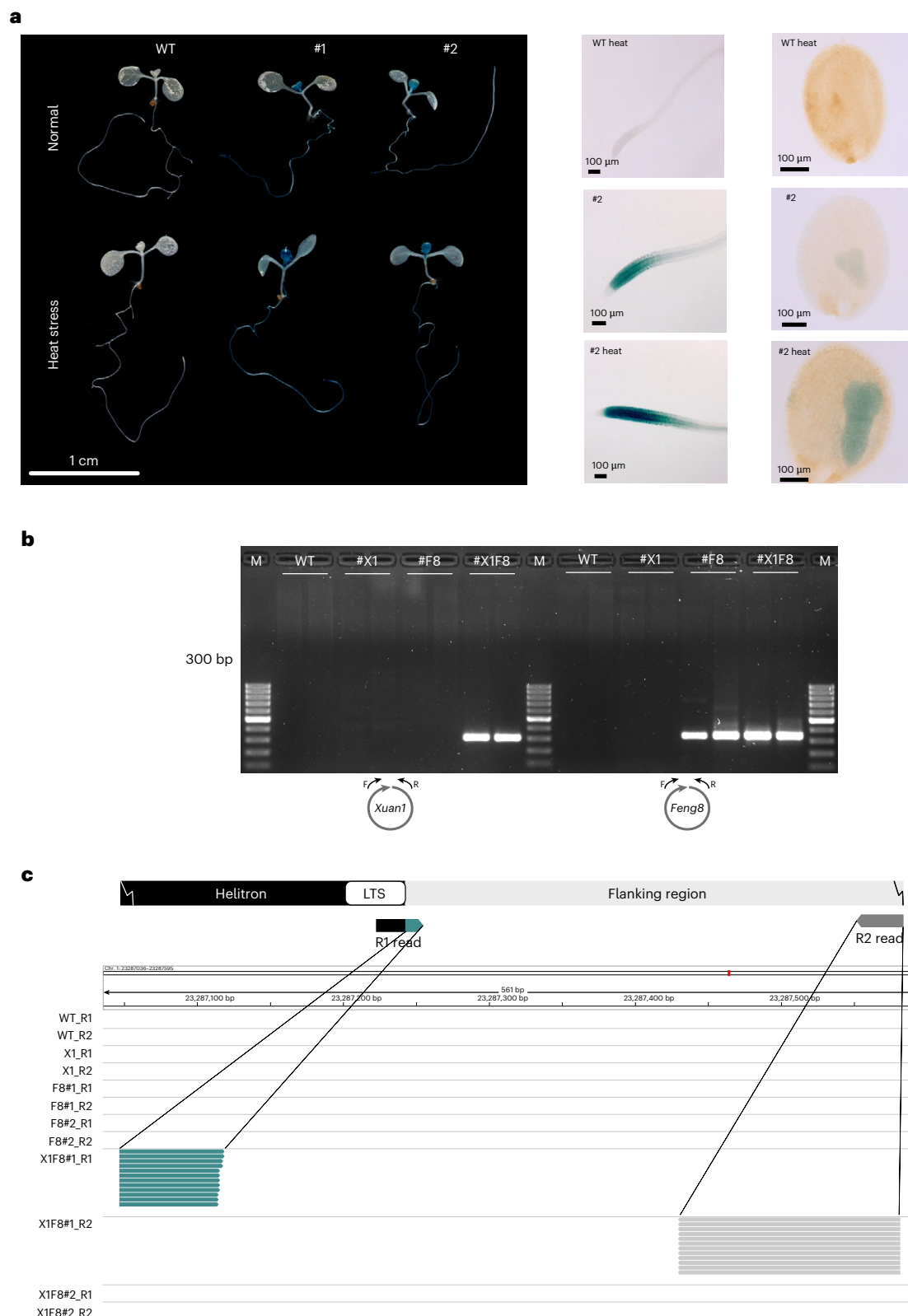


Fig. 3 | Heat-responsive activation of *Feng8* and functional reconstitution of *Xuan-Feng* reactivation in *Arabidopsis*. **a**, 966 bp of the *Feng8* promoter fused to GUS drives basal expression under control conditions (21 °C) and shows markedly enhanced and broader staining following heat treatment (37 °C). Representative whole-seedling images (left) and magnified views of roots and embryos (right) are shown. Two independent lines with three biological replicates were tested. **b**, Detection of *Xuan1* and *Feng8* eccDNA by inverse PCR

in wild type (WT) and transgenic *Arabidopsis* lines expressing *Xuan1* alone (#X1), *Feng8* alone (#F8) or both (#X1F8) after heat treatment. Two independent lines with three biological replicates were tested. The M lanes show the GeneRuler 100-bp DNA ladder. **c**, TED-seq detection of a representative de novo insertion event in *Arabidopsis*. Reads spanning the *Feng8* LTS and flanking genomic DNA are shown. High-confidence paired-end reads (R1 and R2) support insertion junctions in X1F8#1 lines, whereas no insertions are detected in WT or other lines.

only in the #X1F8 line, whereas *Feng8* insertions were detected in both the #F8 and #X1F8 lines. These results indicate that *Xuan1* and *Feng8* can complete their life cycle in a heterologous system, albeit at a very low frequency.

In summary, our study demonstrates that an epigenetically silenced and heat-stress-responsive Helitron can be activated in wheat. This observation is evidenced by transcriptional activity, eccDNA formation and the documentation of novel somatic insertions. Genetic segregation in wheat RILs established a strict association between the presence of *Feng8* and the mobilization of *Xuan1*. This was further supported by heterologous reconstitution in *Arabidopsis*. This demonstrates that *Feng8* is required and sufficient to drive *Xuan1* mobilization, thereby confirming that the *Xuan–Feng* Helitron family relies on *Feng8* as its sole fully intact autonomous element. Together, our findings provide insights into TE biology in wheat and Helitron mobility in general.

Methods

Plant materials and growth conditions

All cultivars of wheat seeds (Arina, CH Nara, Rosatch, CS and Norin61) were obtained from Agroscope GenBank. RILs were obtained from INRAE GenBank. Prior to planting, the wheat seeds were soaked overnight in sterilized water and subsequently sterilized using a 10-min heat shock at 50 °C⁴⁰. Following surface sterilization, three seeds were planted on 10 ml of agar media (0.55%) in a PHCbi MLR-352 growth chamber. The growth chamber was set to long-day conditions, with 16 h of light at 20 °C during the day and 18 °C at night, and a light level of 4 was maintained throughout the experiment. For *Arabidopsis*, seeds were surface sterilized with 70% ethanol and washed three times with sterile water. Sterilized seeds were sown on 1/2 Murashige and Skoog plates, stratified for two days at 4 °C and germinated under long-day conditions (16 h light / 8 h darkness) at 21 °C.

Chemical treatment and stress

The treatment group seeds were cultured on agar media containing two drugs (as suggested by epibreed AG)—namely, 50 µM Juglone (Sigma, H47003) and 40 µM zebularine (Sigma, Z4775) dissolved in DMSO. An equal volume of DMSO (Fisher Scientific, BP231-100) was administered to the control group seeds. Heat stress was induced by exposing six-day-old seedlings to 40 °C for 24 h. For *Arabidopsis*, heat stress was induced by exposing six-day-old seedlings to 37 °C for 24 h.

DNA extraction and eccDNA sequencing

For wheat samples, the uppermost 3-cm section of the stem (coleoptile) was harvested from three individual seedlings, and the coleoptile sheath was carefully removed. The three harvested shoot sections were combined into one tube immediately at the end of the stress period and snap-frozen in liquid nitrogen, followed by storage at –80 °C until DNA extraction. For *Arabidopsis* samples, around 25 whole seedlings were collected for one tube in –80 °C after treatment until DNA extraction. The total DNA was then extracted from the frozen samples using the modified CTAB method⁴¹.

The eccDNA-seq method was employed for all samples, as previously described⁴². Briefly, 2 µg of total DNA was digested with 10 U of PlasmidSafe (LubioScience, E3101K) for 16 h at 37 °C to eliminate linear DNA, followed by enzyme denaturation at 70 °C for 30 min. Next, 100 ng of the digested DNA was precipitated with ethanol supplemented with 1 µl of GlycoBlue (Fisher Scientific, 10391565). Circular DNA was then amplified through RCA using the Illustra TempliPhi kit (GE Healthcare, 25-6400-10) according to the manufacturer's instructions, with the reaction left for 16 h at 30 °C. The RCA reaction was inactivated by heating at 65 °C for 10 min, followed by cooling at 4 °C. The amplified DNA was once again precipitated with ethanol and debranched using 10 U of T7 Endonuclease I (New England Biolabs, M0302S) by incubating at 37 °C for 30 min. After ethanol precipitation, 400 ng of

debranched DNA was used for library preparation with the Nanopore Rapid Barcoding Sequencing Kit (SQK-RBK004). The eccDNA-seq was carried out using PromethION 2 Solo with flow cell FLO-PRO114M, and basecalling was performed using Guppy v.7.1.4 (ref. 43).

Wheat whole-genome sequencing

The high-molecular-weight genomic DNA from the leaf samples was fragmented using a gTube to select fragments of approximately 20 kb. The library preparation was carried out using a ligation sequencing kit (SQK-LSK109), and sequencing was performed on the FLO-PRO002 flow cell of PromethION by Biomarker (BMKGENE) Europe. The clean data obtained from Nanopore sequencing were aligned to the wheat reference genome (*Triticum aestivum*_ArinaLrFor_v3.0, IPK) using minimap2 (ref. 44) v.2.26 with the default parameters. Subsequently, we employed nanomonsv⁴⁵ v.0.7 to detect structural variations.

Whole-genome DNA methylation analysis

High-molecular-weight genomic DNA was extracted from coleoptile tissues and used for library construction. Sequencing libraries were prepared using the Oxford Nanopore Technologies Ligation Sequencing Kit V14 (SQK-LSK114) following the manufacturer's instructions. Libraries were sequenced on an ONT PromethION 2 platform using FLO-PRO114M flow cells, generating approximately 11.15× (wild type) and 12.92× (JZH-treated) genome coverage. Basecalling and modified base detection were performed using Dorado v.0.7.11.2 (ref. 46) with a high-accuracy model enabling simultaneous detection of 5-methylcytosine (5mC) across all sequence contexts. To ensure accurate mapping of *Feng8*-associated reads, the *Feng8* sequence was manually inserted into the ArinaLrFor reference assembly at chromosome 2A: 319212266–319213602. Raw reads were aligned to the modified reference using minimap2 v.2.29 (parameters: -ax lr:hq; -2; -y; -secondary=no). Methylation calls were aggregated using modkit v.0.6.1 with the modkit pileup function, specifying the detection of 5mC and 5hmC. Methylation levels were calculated separately for CG, CHG and CHH contexts using motif-specific parameters (-motif CG, 0; -motif CHG, 0; -motif CHH, 0) and combined strand information. Differential methylation analysis at single-base resolution was performed using the modkit dmr pair function with parameters -base C -fine-grained.

eccDNA loci identification

Basecalled, demultiplexed and trimmed eccDNA-seq reads from the same sample were concatenated. To identify eccDNA formation regions in JZH-treated Arina seedlings, the reads were aligned to the wheat reference genome (*Triticum aestivum*_ArinaLrFor_v3.0, IPK) using ecc_finder v.1.0.0 in map-ont mode with the parameter MAX_SIZE in the Genrich.h dependency script altered to 1,000,000 to enable the processing of long-read alignments.

Inverse PCR to test for eccDNA circle junctions

A total of 100 ng of genomic DNA and 50 ng of RCA-amplified DNA were used as separate templates to generate inverse PCR. Specific primers were designed to span the junction of the *Xuan1* and *Feng8* circles. The PCR reaction was performed in a 20-µl volume according to the manufacturer's protocol of GoTaq G2 DNA Polymerase (Promega, M7841). The PCR amplification conditions consisted of an initial denaturation step at 95 °C for 2 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, extension at 72 °C for 2 min and a final extension step at 72 °C for 10 min. The purification of PCR products was carried out using the Wizard PCR clean-up system (Promega, A9281), following the Sanger sequencing service provided by Microsynth AG.

Helitron annotation and *Xuan–Feng* family identity

For Helitron annotation, we filtered all candidate sequences identified by HELIANO⁴⁷ v.1.2.0 using the parameters -pt 1 -is1 1 -is2 1 -p 1e-5. To identify members of the *Xuan–Feng* family, we used the LTS–RTS of

Xuan1 as the reference LTS–RTS pair file for the HELIANO ts parameter, applying -dis_denovo -nearest-dn 8000 for dis-denovo prediction. The resulting sequences were then manually curated by excluding those containing 'N' bases and selecting 30 bp of the LTS and the RTS to obtain high-confidence members.

Transcript analysis

Total RNA extractions were carried out using the NucleoSpin RNA kit (Macherey-Nagel, 740955.50) for three biological replicate samples under each condition. The extracted RNA samples were subjected to Illumina 150-bp paired-end sequencing using a stranded poly-A directional library at Novogene. Raw mRNA-seq reads were aligned to the complete gene sequences of *Feng1-8* and *Xuan1* using STAR⁴⁸ v.2.7.10a with the parameter outFilterMismatchNmax set to 0 to disallow mismatches. Alignment BAM files were filtered with samtools⁴⁹ v.1.18 to keep only primary mappings and were normalized and converted to bigwig format with the scale function of BAMscale⁵⁰ v.1.0 using the parameters -operation rna -scale custom and -factor with custom normalization factors for each library. Normalization factors were calculated by assembling a de novo reference transcriptome with Trinity⁵¹ v.2.15.1 using all reads and the default parameters, quantifying transcripts with RSEM⁵² v.1.3.3 and bowtie2 (ref. 53) as aligners and applying the function calcNormFactors from the R library edgeR⁵⁴ v.4.0.3 on the posterior mean counts in R v.4.3.2 (ref. 55).

Cloning and sequencing of PCR fragments

Feng8 cloning from Arina and Renan genomic DNA was produced by PCR using the TOPO XL-2 complete PCR cloning kit (Invitrogen, K805010 and K805020). The purified PCR products were ligated with the pCR-XL-2-TOPO vector at a molar ratio of 1:1 (transcript:vector). The ligation reaction was carried out at 25 °C for 30 min. Subsequently, 2 µl of the ligation product was transformed into One Shot OmniMAX 2 T1^R chemically competent *Escherichia coli* using a heat-shock method. The transformed cells were plated on selective agar plates, and individual colonies were selected for further analysis. The presence of the desired inserts in the colonies was confirmed by PCR, and the plasmids were subsequently purified using the PureLink quick plasmid miniprep kit (Invitrogen, K210010). To ensure the integrity of the cloned inserts, the entire plasmid sequence was determined using Microsynth AG.

Transposase analyses and structure prediction

The 3D structure prediction of the protein was obtained using AlphaFold³⁴ v.2.3.1. Five parallel model predictions were conducted for each sequence, and the structures with the highest predicted local distance difference test scores were selected for analysis and visualization using PyMOL v.2.1 (ref. 56).

Phylogenetic tree construction

To generate the maximum-likelihood phylogenetic tree for transposase, the amino acid sequence of FENG8 was employed as the query for a blastp search against the NCBI database. Top-scoring hits from 83 different species were chosen to construct the phylogenetic tree. Multiple sequence alignment was carried out using MUSCLE⁵⁷ v.5.1, and the alignment was subsequently converted to phylip format using trimAl⁵⁸ v.1.4. The phylogenetic tree was constructed using IQ-TREE⁵⁹ v.2.12 with a bootstrap value of 1,000 replicates.

Selection of 21 CS × Renan RILs based on SNP genotyping data

We analysed SNP genotyping data (Axiom array TaBW420k) available for 282 individuals originating from a cross between CS and Renan⁶⁰, to select 21 individuals that carry either CS or Renan alleles at both *Feng8* (chromosome 2A) and *Xuan1* (chromosome 6D) loci. The positions of the markers were identified by mapping SNP context sequences on the CS IWGSC RefSeq v.2.1 using the pipeline nf_remapMarkers (https://forgemia.inra.fr/umr-gdec/nf_remapmarkers). Although

Feng8 is absent from CS, its insertion site is at position 323357927 on chromosome 2A. The position of *Xuan1* is 169657404 on chromosome 6D. We retrieved SNP alleles in a window of 100 Mb centred on each of the two loci on chromosomes 2A and 6D. This represented 496 and 163 SNPs on these two chromosomes, respectively. We kept only 132 individuals, all alleles within 100 Mb, encompassing each locus derived from the same parent. We classified the 132 individuals into four haplotypes: A/A, A/B, B/B and B/A (allele A is from CS, and allele B is from Renan; *Feng8/Xuan1*). B/B and B/A individuals were expected to carry a functional *Feng8* originating from the Renan parent, while *Feng8* was predicted to be absent for A/B and A/A individuals. We eventually randomly selected 21 individuals for their ability to produce eccDNAs so that 5, 5, 5 and 6 were of the following haplotypes: B/B (#2038, #2483, #2883, #2986 and #2998), B/A (#2815, #2885, #2894, #2916 and #3091), A/B (#3101, #2620, #2866, #2893 and #3060) and A/A (#2968, #2726, #2825, #2858, #2952 and #2543), respectively.

Modified PCIP-seq to capture de novo Helitron insertions

The method has been described in detail in Tang et al.³⁸. The single-guide RNAs (sgRNAs) and primers were designed on the basis of the *Xuan1* and *Feng8* sequences. The experimental protocols and data analysis procedures were identical for *Xuan1* and *Feng8*, except for the initial cleavage: one sgRNA (1,357 bp from the LTS) targeted *Xuan1*, while two sgRNAs were used to cut two sites in *Feng8* (1,222 bp from the LTS and 1,068 bp from the RTS). The description below uses *Xuan1* as an example.

Molecular biology. Using 500 ng of genomic DNA as starting material, fragments containing *Xuan1* sequences were cleaved using sgRNAs (Integrated DNA Technologies) targeting sequences at 1,357 bp from the LTS and *S. pyogenes* Cas9 (New England Biolabs, M0386S). The digested DNA was further mechanically sheared to ~3 kb using a Megaruptor-1 (Diagenode), end-repaired using a NEBNext EndRepair Module (New England Biolabs, E6050L) and purified with Agencourt AMPure XP beads (Beckman Coulter, A63881). The resulting DNA fragments were circularized using T4 DNA Ligase (New England Biolabs, M0202L), residual linear fragments were eliminated with Plasmid-Safe-ATP-Dependent DNase (Epicentre, E3110K), the remaining fragments were purified and reaction products were split in two aliquots. Circular molecules encompassing *Xuan1* sequences were then reopened with *S. pyogenes* Cas9 using distinct sgRNAs for the two and purified. Encompassing linear fragments were inverse-PCR-amplified using aliquot-specific primer pairs with LongAmp Taq DNA Polymerase (New England Biolabs, M0533S) and purified. The purified amplicons were mechanically sheared to ~350 bp using a Bioruptor-pico (Diagenode), sequencing libraries were generated using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, E7645L) and indexed libraries were pooled and sequenced on a Novaseq 6000 sequencer (Illumina) targeting ~8 million 150-bp paired-end reads per library.

Data processing. The ensuing sequence reads were demultiplexed, the reads were quality assessed using fastQC⁶¹, adapter sequences were trimmed using Cutadapt⁶² and trimmed reads were mapped to the CS IWGSC RefSeq genome⁵ using BWA-MEM⁶³ and converted to BAM format using SAMtools⁶⁴. Using a custom-made Python script, we first identified clipped reads using CIGAR information. We selected clipped reads with mapping quality ≥40 and a minimum of 10 clipped bases. We then mapped the clipped reads to the segments of the corresponding *Xuan1* genome. We specified an alignment score ≥0.6 to declare a hit and labelled the read as either an insertion site (IS) or a shearing site (SS) read. When possible, we extended the alignment with *Xuan1* into non-clipped bases to refine the positions of the SS and IS. We then merged SS and IS sites with the same 'breakpoint', thereby identifying candidate SS and IS supported by multiple concordant reads. We then

paired the IS with their cognate SS. The pairing was based on orientation (for example, a 5' SS should be located upstream of the 5' IS for a *Xuan1* element in sense orientation, and downstream of the 5' IS for a *Xuan1* in antisense orientation) and distance (the maximum distance between IS and SS was set at 5 kb).

Vector construction and *Arabidopsis* transformation

To generate the GUS reporter lines, the first 966-bp fragment of *Feng8* before the translation start site was amplified and used as the putative promoter sequence. The fragment was cloned upstream of the GUS reporter gene in a modified pGWB553 vector. The *Xuan1* element was amplified and inserted into the pGAPBRHyg vector at the *SpeI* restriction site without an additional promoter sequence. The full-length *Feng8* sequence was introduced into the pBGW0 binary vector using Gateway recombination (Invitrogen) according to the manufacturer's instructions. Subsequently, the *Xuan1* element was inserted into the pBGW0-*Feng8* construct at the *SbfI* site using the In-Fusion HD Cloning system (Takara) to generate the pBGW0-*Xuan1_Feng8* construct. No additional promoter sequence was included in this construct. All constructs were verified by Sanger sequencing prior to transformation. For plant transformation, constructs were introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation. *Arabidopsis* (Col-0 ecotype) plants were transformed using the floral dip method. Transgenic plants were selected on Murashige and Skoog medium supplemented with the appropriate antibiotics.

GUS staining assay

Six-day-old seedlings were divided into control and heat-treatment groups. Control seedlings were maintained at 21 °C, whereas heat-treated seedlings were incubated at 37 °C for 12 h under identical continuous light conditions. Following treatment, seedlings were subjected to histochemical GUS staining. The samples were incubated in GUS staining buffer containing 50 mM sodium phosphate buffer (pH 7.0), 10 mM EDTA, 0.1% (v/v) Triton X-100, 1 mM potassium ferrocyanide ($K_4Fe(CN)_6$), 1 mM potassium ferricyanide ($K_3Fe(CN)_6$) and 1 mg ml⁻¹ 5-bromo-4-chloro-3-indolyl β-D-glucuronic acid (X-Gluc). The seedlings were incubated at 20 °C for 4 h. After staining, tissues were cleared in 70% (v/v) ethanol until chlorophyll was fully removed. For seed analysis, seeds were mounted in Hoyer's solution on microscope slides and incubated at room temperature for four days prior to imaging.

TEd-seq library construction and analysis

TEd-seq libraries were constructed with modifications from Vendrell-Mir et al.³⁹. Briefly, 1 µg genomic DNA was sheared to 250–500 bp using a Bioruptor Plus (three cycles of 15 s on, 30 s off). Fragmented DNA was end-repaired and A-tailed using the NEBNext Ultra II End Prep Enzyme Mix, followed by ligation of TEd-seq forked adapters. Adapter-ligated DNA was size-selected with AMPure XP beads and subjected to selective amplification of TE-containing fragments using NEBNext Ultra II Q5 Master Mix and outer primers (TEDseq_outer_for and TE_outer_rev). PCR products were purified and size-selected (0.9× beads), followed by a second round of nested PCR incorporating custom Illumina P5 and indexed P7 adapters. Final libraries were purified (0.9× beads), eluted in 10 mM Tris-HCl and sequenced (paired-end, 300 cycles) on a NextSeq 1000 platform.

Paired-end reads were analysed using a custom pipeline to identify genomic Helitron insertion sites. In this design, Read 1 (R1) corresponds to the TE-adjacent end, whereas Read 2 (R2) represents the paired genomic flank. Reads were first filtered using TE-specific primers (24 bp) with full-length exact matching in cutadapt (v.5.2). Primer-positive reads were subsequently filtered using TE extremity sequences (10 bp for *Xuan1* and 14 bp for *Feng8*), also requiring full-length exact matches. Extremity sequences were trimmed, and only read pairs retaining ≥15 bp of genomic sequence were retained.

Trimmed R1 reads (R1_ext) were mapped to the TAIR10 reference genome using BWA-MEM (v.0.7.18). Only fully aligned reads without soft or hard clipping were considered. The genomic alignment start position of each R1_ext read was defined as the insertion coordinate. Trimmed R2 reads were independently mapped, and only fully aligned reads were retained. R1-derived insertion candidates were required to have R2 support within ±400 bp to ensure paired-end consistency. Insertions overlapping within 1 kb across samples were excluded from the analyses.

All primers used for this study are listed in Supplementary Table 5.

Reporting summary

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

Data availability

The *Xuan1* and *Feng8* sequences have been submitted to NCBI GenBank with accession numbers [PP376094](#) and [PP376095](#). The sequencing data from this study have been submitted to the European Nucleotide Archive (www.ebi.ac.uk/ena/, accession number [ERP157467](#)) under project number [PRJEB72688](#). The raw whole-genome sequencing reads are under fastq accessions [ERR12723706–ERR12723708](#), and the raw eccDNA-Seq reads for Arina are under the accessions [ERR12724408–ERR12724413](#). The raw RNA-seq reads are under the accessions [ERR12724677–ERR12724682](#). Positions of TaBW420K SNP markers on CS IWGSC RefSeq v.2.1 are available at <https://doi.org/10.57745/YXNWZK>. The raw fastq files for the PCIP-seq experiments applied to the 24 RILs have been submitted to the European Nucleotide Archive under project number [PRJEB75199](#). The raw fastq files for the TEd-seq experiments applied to *Arabidopsis* lines have been submitted to the European Nucleotide Archive under project number [PRJEB108718](#). Source data are provided with this paper.

Code availability

The pipeline for identifying genomic Helitron insertion sites is available without restriction via GitHub at https://github.com/Lijingtangbo/TE-rate_manuscript.

References

- Liu, P., Cuerda-Gil, D., Shahid, S. & Slotkin, R. K. The epigenetic control of the transposable element life cycle in plant genomes and beyond. *Annu. Rev. Genet.* **56**, 63–87 (2022).
- Kapitonov, V. V. & Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **98**, 8714–8719 (2001).
- Barro-Trastoy, D. & Köhler, C. Helitrons: genomic parasites that generate developmental novelties. *Trends Genet.* **40**, 437–448 (2024).
- Vitte, C., Fustier, M. A., Alix, K. & Tenaillon, M. I. The bright side of transposons in crop evolution. *Brief. Funct. Genom. Proteom.* **13**, 276–295 (2014).
- The International Wheat Genome Sequencing Consortium (IWGSC) et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018).
- Naito, K. et al. Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl Acad. Sci. USA* **103**, 17620–17625 (2006).
- Wicker, T. et al. Transposable element populations shed light on the evolutionary history of wheat and the complex co-evolution of autonomous and non-autonomous retrotransposons. *Adv. Genet.* **3**, 2100022 (2021).
- Klein, S. P. & Anderson, S. N. The evolution and function of transposons in epigenetic regulation in response to the environment. *Curr. Opin. Plant Biol.* **69**, 102277 (2022).

9. Gebrie, A. Transposable elements as essential elements in the control of gene expression. *Mob. DNA* **14**, 9 (2023).
10. McClintock, B. Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216 (1956).
11. Thomas, J. & Pritham, E. J. Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol. Spectr.* **3**, MDNA3-0049-2014 (2015).
12. Wells, J. N. & Feschotte, C. A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).
13. Xiong, W., He, L., Lai, J., Dooner, H. K. & Du, C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc. Natl Acad. Sci. USA* **111**, 10263–10268 (2014).
14. Yang, L. & Bennetzen, J. L. Structure-based discovery and description of plant and animal Helitrons. *Proc. Natl Acad. Sci. USA* **106**, 12832–12837 (2009).
15. Yang, L. & Bennetzen, J. L. Distribution, diversity, evolution, and survival of Helitrons in the maize genome. *Proc. Natl Acad. Sci. USA* **106**, 19922–19927 (2009).
16. Li, S.-F. et al. Landscape and evolutionary dynamics of Helitron transposons in plant genomes as well as construction of online database HelDB. *J. Syst. Evol.* **61**, 919–931 (2023).
17. Borgognone, A., Castanera, R., Muguerza, E., Pisabarro, A. G. & Ramírez, L. Somatic transposition and meiotically driven elimination of an active helitron family in *Pleurotus ostreatus*. *DNA Res.* **24**, 103–115 (2017).
18. Mirouze, M. et al. Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461**, 427–430 (2009).
19. Ito, H. et al. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**, 115–120 (2011).
20. Lisch, D. Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* **60**, 43–66 (2009).
21. Grabundzija, I. et al. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat. Commun.* **7**, 10716 (2016).
22. Grabundzija, I., Hickman, A. B. & Dyda, F. Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. *Nat. Commun.* **9**, 1278 (2018).
23. Thieme, M. et al. Inhibition of RNA polymerase II allows controlled mobilisation of retrotransposons for plant breeding. *Genome Biol.* **18**, 134 (2017).
24. Chao, S.-H., Greenleaf, A. L. & Price, D. H. Juglone, an inhibitor of the peptidyl-prolyl isomerase Pin1, also directly blocks transcription. *Nucleic Acids Res.* **29**, 767–773 (2001).
25. Poborilova, Z. et al. DNA hypomethylation concomitant with the overproduction of ROS induced by naphthoquinone juglone on tobacco BY-2 suspension cells. *Environ. Exp. Bot.* **113**, 28–39 (2015).
26. Zhou, L. et al. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J. Mol. Biol.* **321**, 591–599 (2002).
27. Finnegan, E. J. et al. Zebularine treatment is associated with deletion of FT-B1 leading to an increase in spikelet number in bread wheat. *Plant Cell Environ.* **41**, 1346–1360 (2018).
28. Horváth, V., Merenciano, M. & González, J. Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet.* **33**, 832–841 (2017).
29. Zhang, P., Peng, H., Llauro, C., Bucher, E. & Mirouze, M. ecc_finder: a robust and accurate tool for detecting extrachromosomal circular DNA from sequencing data. *Front. Plant Sci.* **12**, 743742 (2021).
30. Peng, H., Mirouze, M. & Bucher, E. Extrachromosomal circular DNA: a neglected nucleic acid molecule in plants. *Curr. Opin. Plant Biol.* **69**, 102263 (2022).
31. Walkowiak, S. et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**, 277–283 (2020).
32. Cavrak, V. V. et al. How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet.* **10**, e1004115 (2014).
33. Kosek, D. et al. The large bat Helitron DNA transposase forms a compact monomeric assembly that buries and protects its covalently bound 5'-transposon end. *Mol. Cell* **81**, 4271–4286.e4 (2021).
34. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
35. Zhu, T. et al. Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant J.* **107**, 303–314 (2021).
36. Aury, J.-M. et al. Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. *GigaScience* **11**, giac034 (2022).
37. Artesi, M. et al. PCIP-seq: simultaneous sequencing of integrated viral genomes and their insertion sites with long reads. *Genome Biol.* **22**, 97 (2021).
38. Tang, L. et al. GWAS reveals determinants of mobilization rate and dynamics of an active endogenous retrovirus of cattle. *Nat. Commun.* **15**, 2154 (2024).
39. Vendrell-Mir, P., Leduque, B. & Quadrana, L. Ultra-sensitive detection of transposon insertions across multiple families by transposable element display sequencing. *Genome Biol.* **26**, 48 (2025).
40. Watts, J. E., de Villiers, O. T. & Watts, L. Sterilization of wheat seeds for tissue culture purposes. *S. Afr. J. Bot.* **59**, 641–642 (1993).
41. Ferdous, J. A quick DNA extraction protocol: without liquid nitrogen in ambient temperature. *Afr. J. Biotechnol.* **11**, 6956–6964 (2012).
42. Roquis, D. et al. Genomic impact of stress-induced transposable element mobility in *Arabidopsis*. *Nucleic Acids Res.* **49**, 10431–10447 (2021).
43. Guppy v.7.1.4 (Oxford Nanopore Technologies, 2023); <https://nanoporetech.com/software/other/guppy>
44. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
45. Shiraishi, Y. et al. Precise characterization of somatic complex structural variations from tumor/control paired long-read sequencing data with nanomonsv. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkad526> (2023).
46. Dorado v.0.7.11.2 (Oxford Nanopore Technologies, 2025); <https://github.com/nanoporetech/dorado>
47. Li, Z., Gilbert, C., Peng, H. & Pollet, N. Discovery of numerous novel Helitron-like elements in eukaryote genomes using HELIANO. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkae679> (2024).
48. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
49. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
50. Pongor, L. S. et al. BAMscale: quantification of next-generation sequencing peaks and generation of scaled coverage tracks. *Epigenetics Chromatin* **13**, 21 (2020).
51. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
52. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

54. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
55. R Core Team. *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/> (R Foundation for Statistical Computing, 2025).
56. PyMOL v.2.1 (Schrödinger, 2018).
57. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
58. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
59. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
60. Darrier, B. et al. High-resolution mapping of crossover events in the hexaploid wheat genome suggests a universal recombination mechanism. *Genetics* **206**, 1373–1388 (2017).
61. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* (Babraham Bioinformatics, 2010); <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
62. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
63. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://doi.org/10.48550/arXiv.1303.3997> (2013).
64. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

Acknowledgements

We acknowledge the Mésocentre Clermont-Auvergne and the AuBi platform of Université Clermont Auvergne for providing computational and storage resources. We thank the GIGA Genomics platform for generating the next-generation sequencing data used for de novo insertion detection. We thank J. Zhao for advice on *k*-mer-based analysis of insertion polymorphisms in the wheat population. We thank the lab of R. Ulm for providing plants, plasmids and constructive insights. We thank A. Valentin and the greenhouse facility of the Max Planck Institute for technical assistance.

Author contributions

H.P., L.T., N.H. and R.K. conducted the experiments. H.P., R.K., H.R., S.M., A.B.H., F.D., C.C., F.C. and E.B. analysed the data. D.F., P.S. and F.C. provided the materials. H.P., F.C., C.K. and E.B. designed the experiments and wrote the paper.

Funding

This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation

programme (725701, BUNGEE, to E.B.); the Agence Nationale de la Recherche and Fonds National Suisse (ANR-21-PRCI-CE02, FNS 310030E_205554 'CropCircle' to E.B.); the China Scholarship Council Grant (201806990012 to H.P.); the Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, USA (to F.D. and A.B.H.); the Fund for Scientific Research—FNRS (PDR grant to C.C., TE-rate project), Belgium; and the Max Planck Society, Germany.

Competing interests

E.B. is a member of the board of epibreed AG, a spin-off company involved in plant breeding. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41477-026-02319-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41477-026-02319-3>.

Correspondence and requests for materials should be addressed to Etienne Bucher.

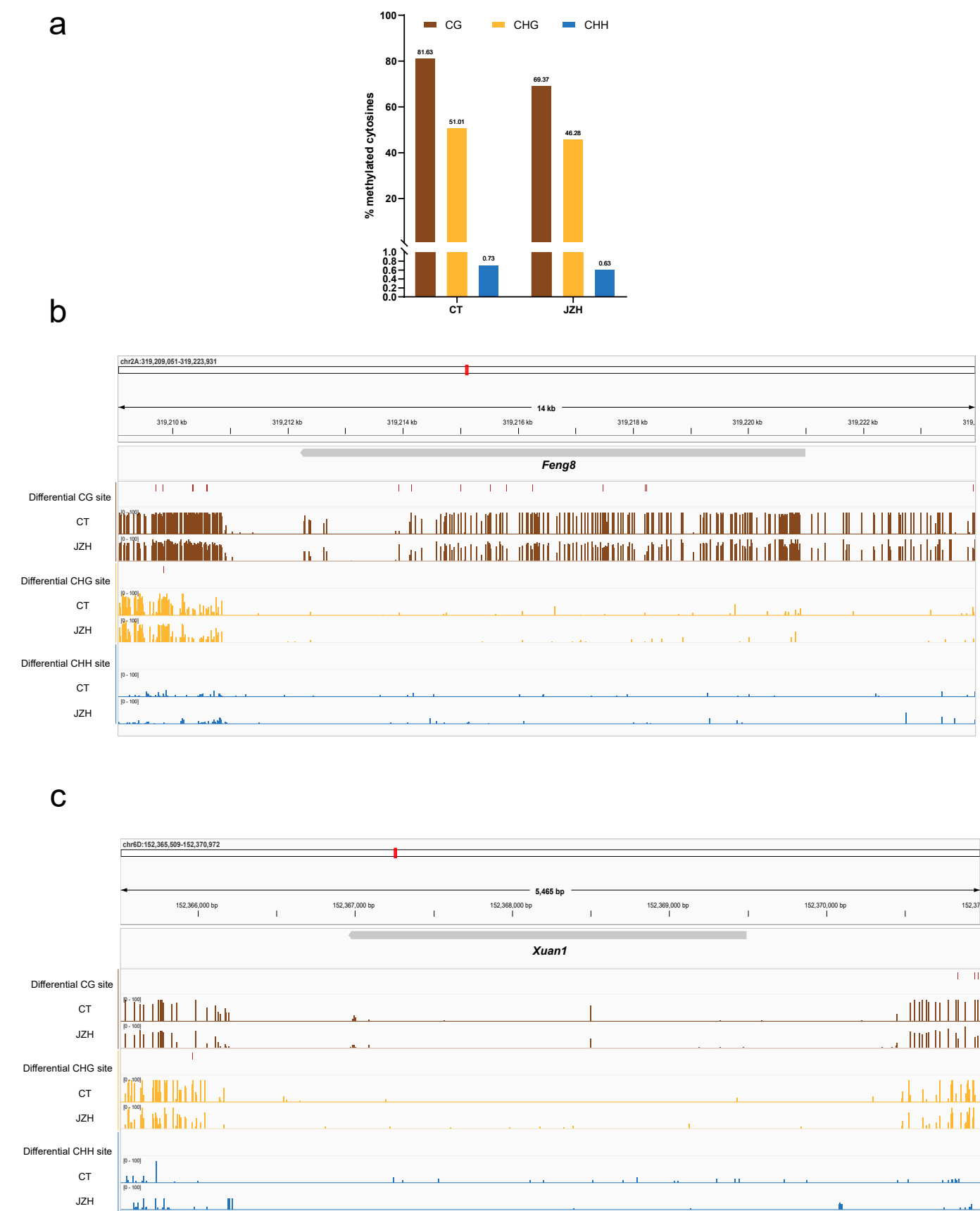
Peer review information *Nature Plants* thanks Jungnam Cho, Weiqiang Qian and Leandro Quadrana for their contribution to the peer review of this work.

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.

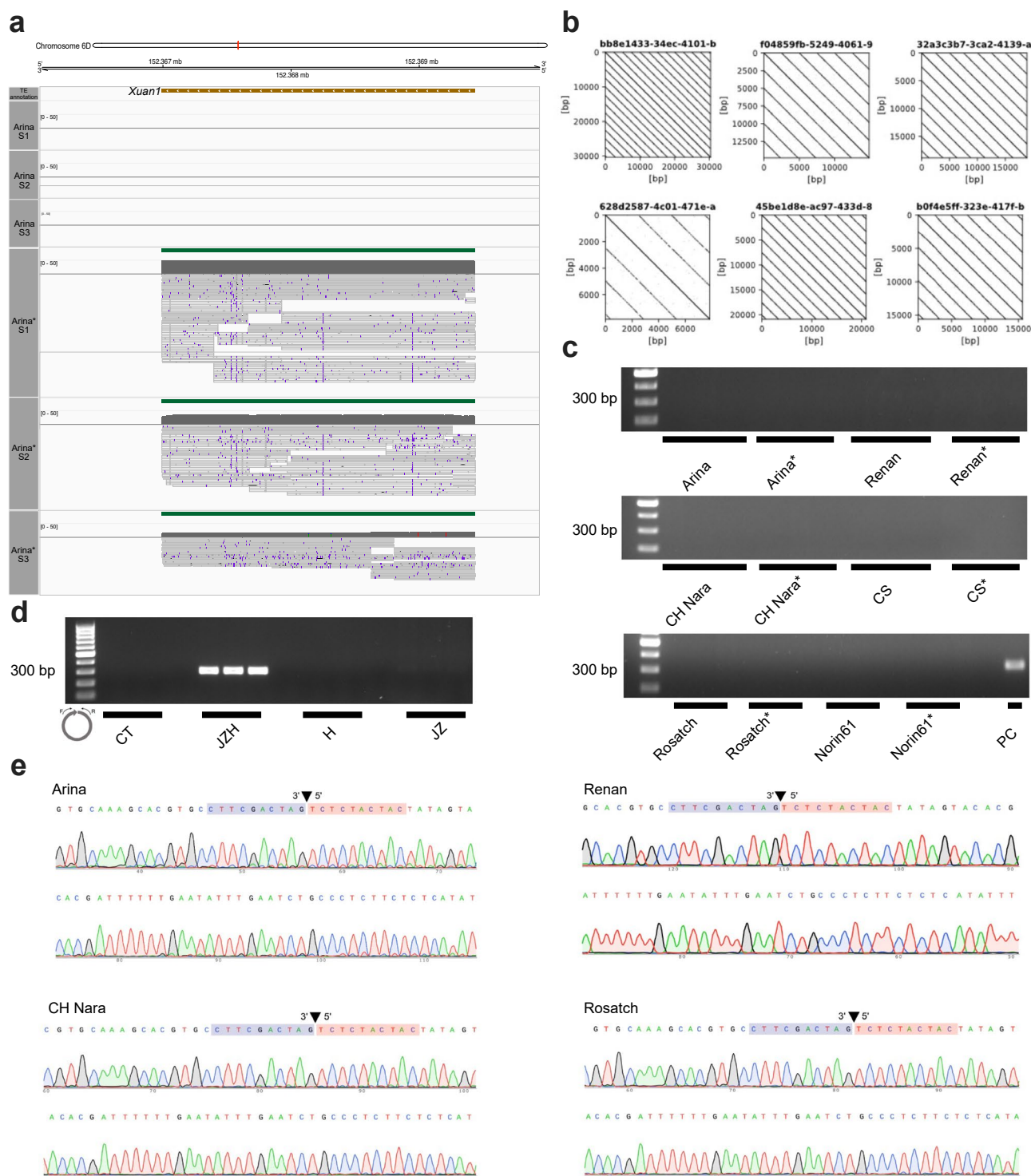
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026



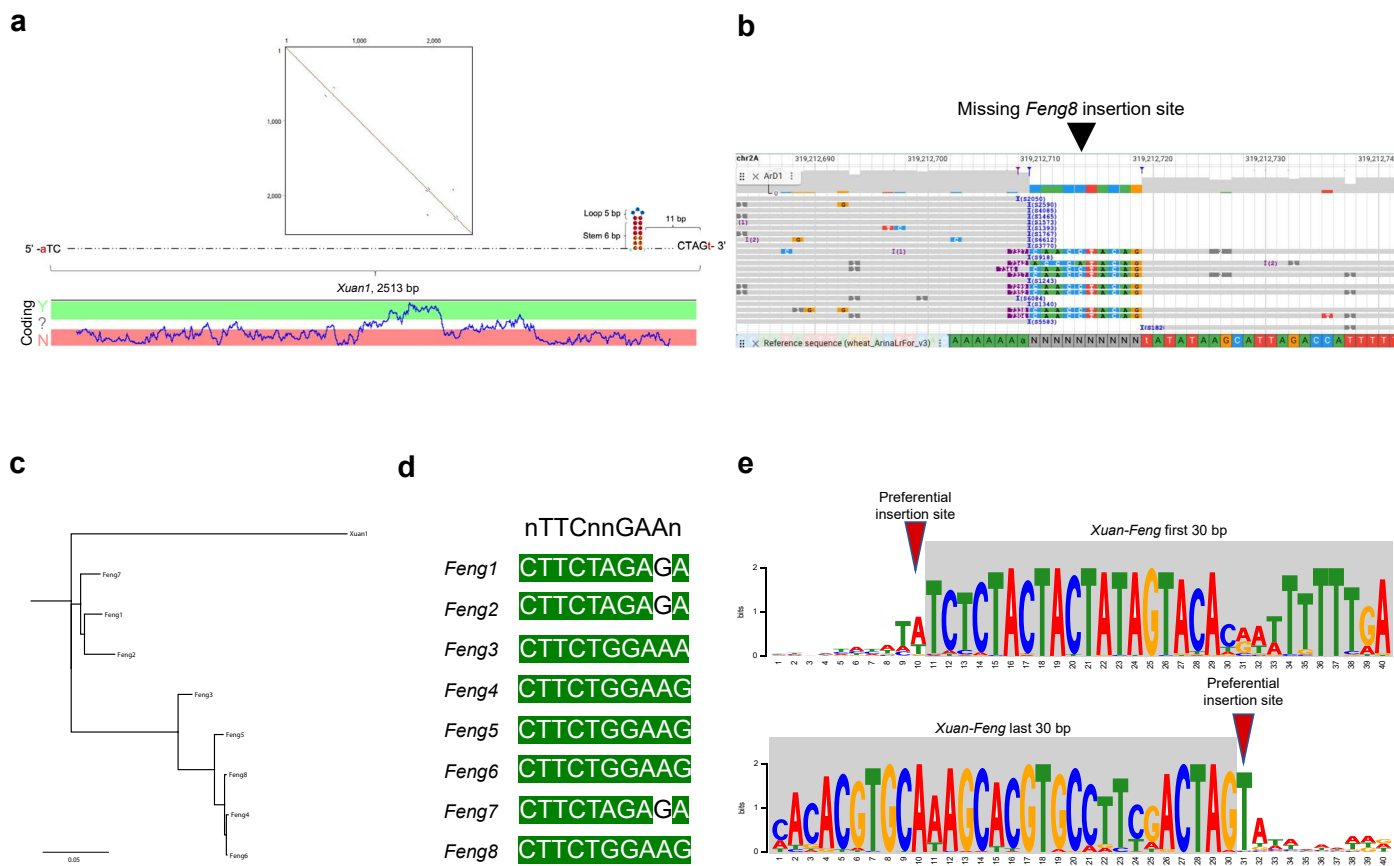
Extended Data Fig. 1 | Genome-wide and locus-specific DNA methylation changes following JZH treatment. a, Global levels of cytosine methylation across the three sequence contexts (CG, CHG, and CHH) in untreated and JZH-treated plants. **b,** Integrative Genomics Viewer (IGV) visualization of cytosine

methylation in the *Feng8* locus. Tracks represent CG (brown), CHG (yellow), and CHH (blue) methylation. Significantly differentially methylated cytosine positions are highlighted in red. **c,** IGV visualization of cytosine methylation in the *Xuan1* locus.

**Extended Data Fig. 2 | Identification and verification of *Helitron* eccDNA.**

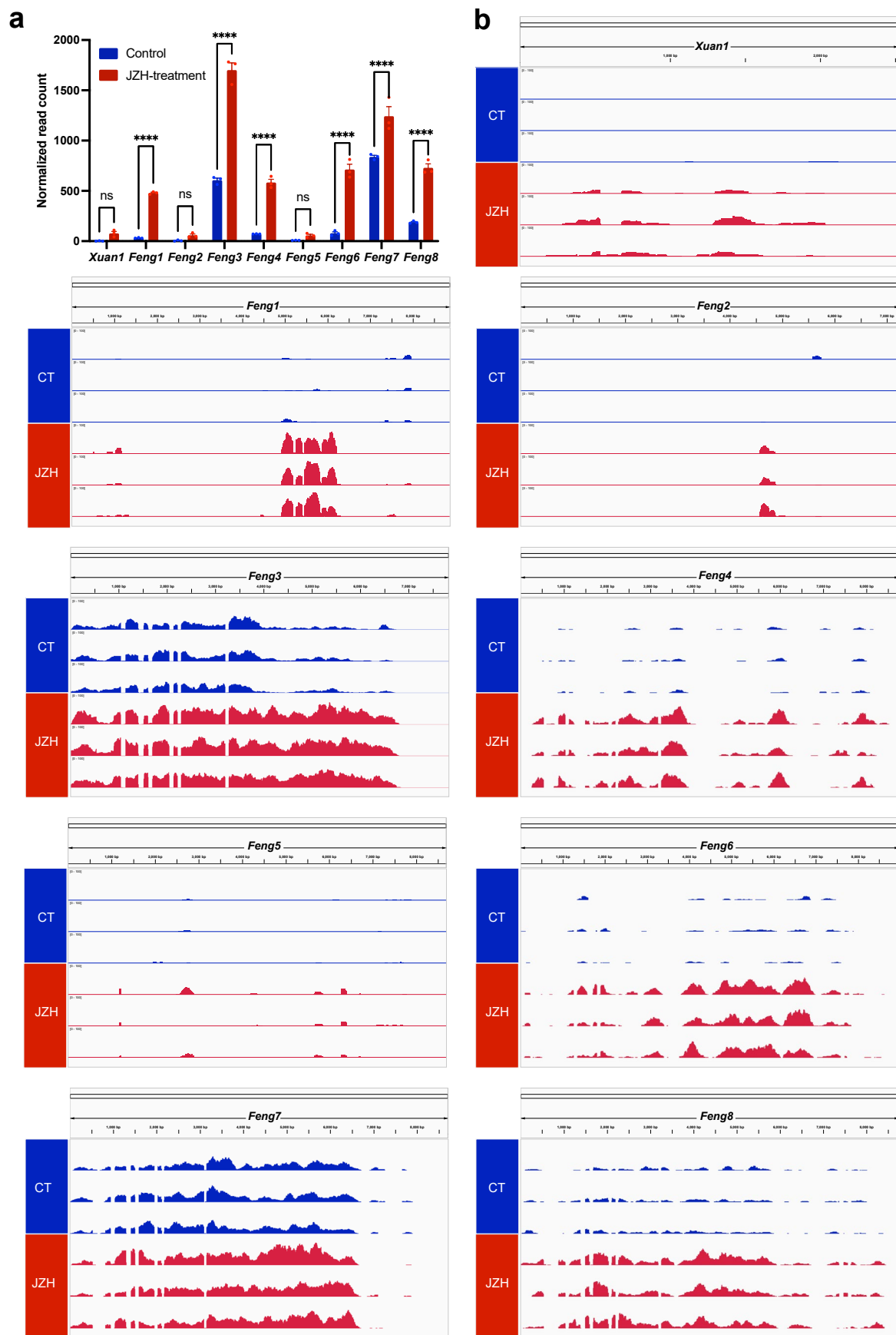
a, Nanopore eccDNA sequencing analysis of untreated and treated *Arina* seedlings. The chromosomal position is shown on the top, the first row presents our *Helitron* annotation of the *ArinaLrFor* genome³¹. Each row below the TE annotation represents an individual sample with three biological replicates from the control and JZH treated wheat seedlings. The green lines depict the eccDNA loci that were detected by *ecc_finder*²⁹, the dark gray oblong shape indicates normalized read coverage and below are the individual Nanopore reads mapped to this region (gray reads are single-mapped and white reads mapped to multiple loci in the genome). **b**, Dot-plot analyses visualizing the self-similarity of selected single long reads containing tandem repeats of the *Helitron* in JZH-treated *Arina*. **c**, Confirmation of absence of genomic *Xuan1* tandem repeats in the corresponding wheat varieties using inverse PCR with genomic DNA as template

(no RCA), the primer localization depicted on the left (F, R). The first lane shows the GeneRuler 100 bp DNA ladder, each cultivar was tested using three biological replicates issued from the control and JZH-treated groups. RCA eccDNA from treated *Arina* plants was used as a template for the positive controls (PC). *Xuan1* eccDNA can only be detected following RCA. The asterisk (*) indicates JZH-treated samples. **d**, Detection of *Helitron* eccDNA in the *Arina* variety using inverse PCR on RCA-amplified eccDNA; the primer localization depicted on the left (F, R). The first lane shows the GeneRuler 100 bp DNA ladder; each treatment was tested with three biological replicates issued from the control (CT), JZH treatment, heat treatment (H), and JZ-treatment groups. **e**, Chromatograms from Sanger sequencing, the blue background highlights the 10 bp of *Helitron* RTS, and the red background highlights the 10 bp of *Helitron* LTS, and the inverted black triangle indicates the circle junction.



Extended Data Fig. 3 | Sequence features of the *Xuan-Feng Helitron* family. **a**, Self-dot plot of the *Xuan1* sequence and schematic representation of the molecular features. The graph below indicates the protein coding capacity of the sequence, with red indicating low and green indicating high. The hairpin structure is depicted as colored rings. **b**, Genome browser view of the *Feng8* locus on chr2A:319,212,266 of the *ArinaLrFor* reference genome where the *de novo* assembly failed to assemble *Feng8* (as indicated by the multiple N's in the reference sequence shown on the bottom). The gray horizontal bars represent individual nanopore sequencing reads and insertions (numbers with magenta

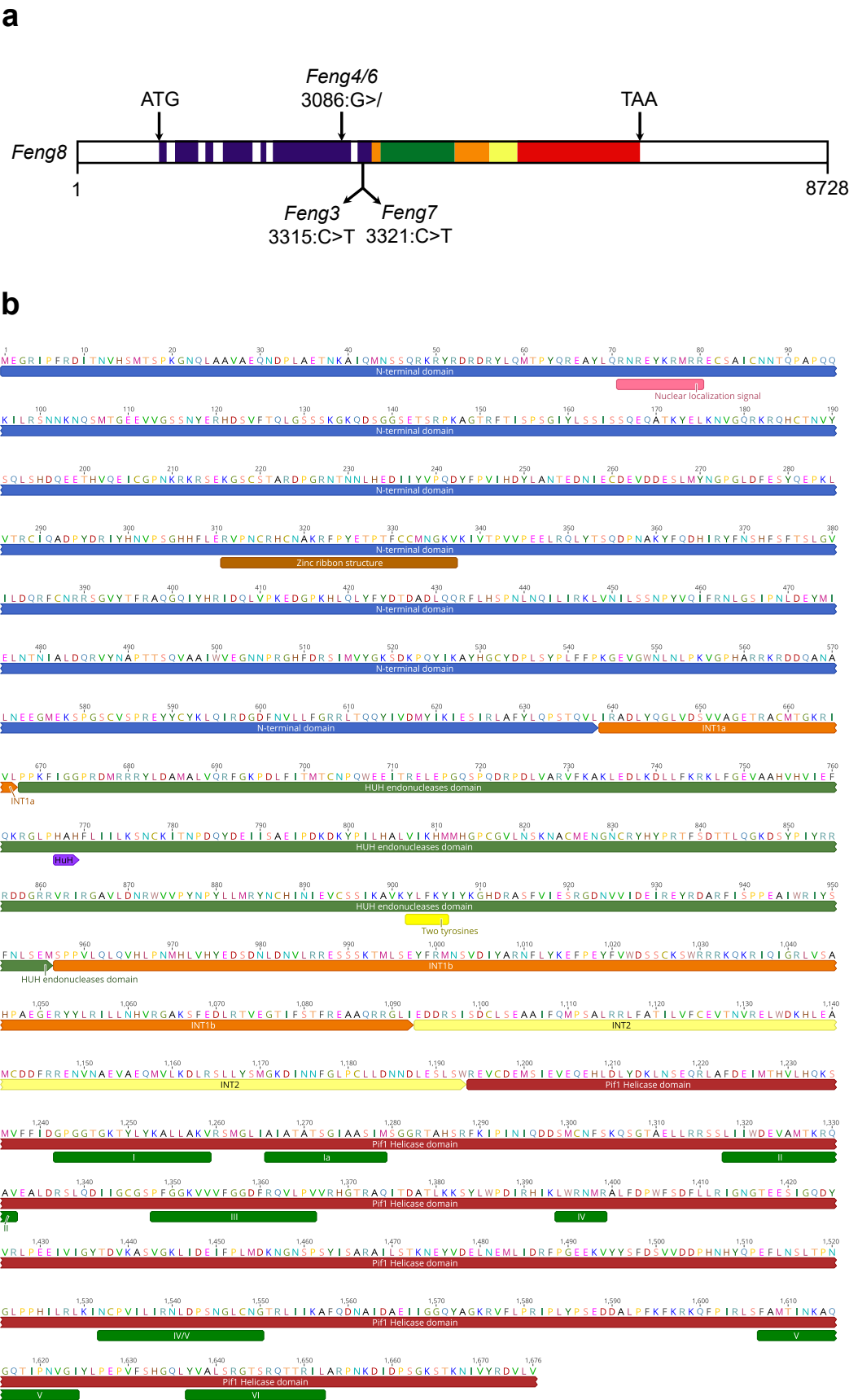
backgrounds) and mismatches (colored bases) compared to the reference sequence. **c**, Phylogenetic tree depicting the relationship between *Feng* members, *Xuan1* was included as an outlier. **d**, Comparison of the predicted heat response element (HRE) motif in the eight *Feng* members. The consensus bases are highlighted. **e**, Sequence motif logo illustrating the 40 bp boundaries of the *Xuan-Feng* family members, including 10 bp upstream and downstream sequences for each 5' and 3' ends. The gray background highlights the *Helitron* sequence.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Transcript levels of *Xuan1* and *Feng1-8* assessed by RNA-seq. a, Quantification of the RNA-seq data obtained from control (blue) and JZH- treated (red) Arina seedlings. Three biological replicates derived from independent plants were analyzed. The statistics use two-way ANOVA of normalized read count mean with SEM; the stars indicate the adjusted *P*

value < 0.0001. The exact *P* values are provided in Source Data Extended Data Fig. 4. **b**, Genome browser views showing three biological replicates of normalized transcript levels assessed by RNA-seq for *Xuan1* and the eight *Feng* sequences under control conditions (blue) and resulting from JZH treatment (red).

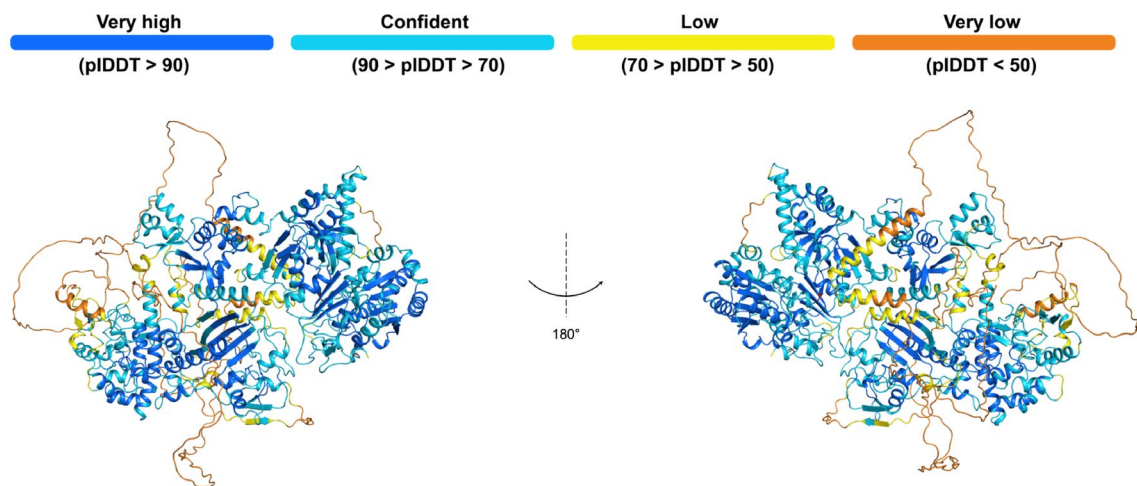
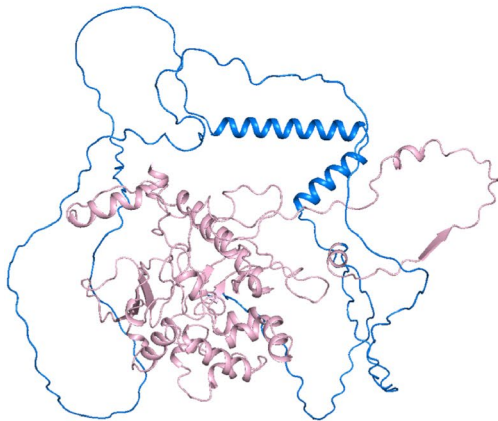
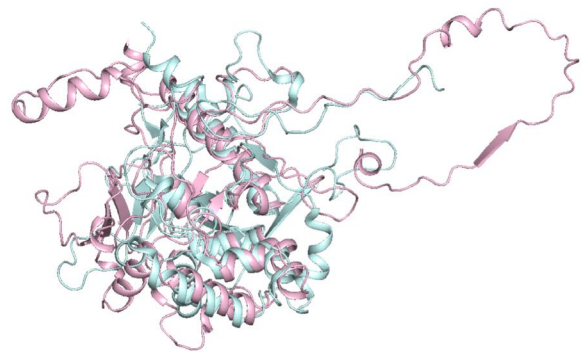
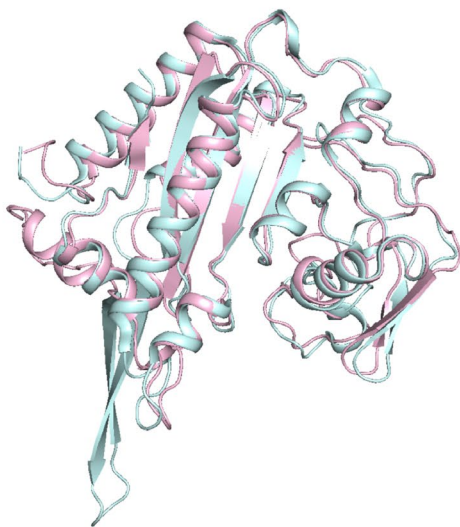
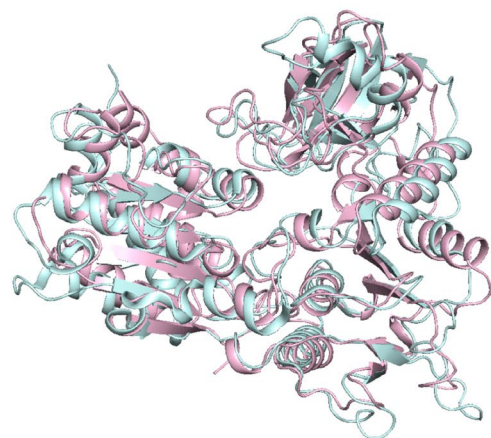


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Graphical representation of the FENG8 protein sequence highlighting the conserved domains of the transposase.

a, Schematic diagram depicting mutations that impact the protein coding sequences in *Feng3* (nonsense mutation), *Feng4* (frameshift mutation), *Feng6* (frameshift mutation), and *Feng7* (frameshift mutation) represented on the

intact *Feng8* sequence. The colors on the exons correspond to the different transposase domains shown in Fig. 2. **b**, Residues involved in the zinc ribbon are highlighted in brown, the HUH motif is highlighted in purple, the Y2 motif in yellow, and the eight Helicase motifs are indicated in green.

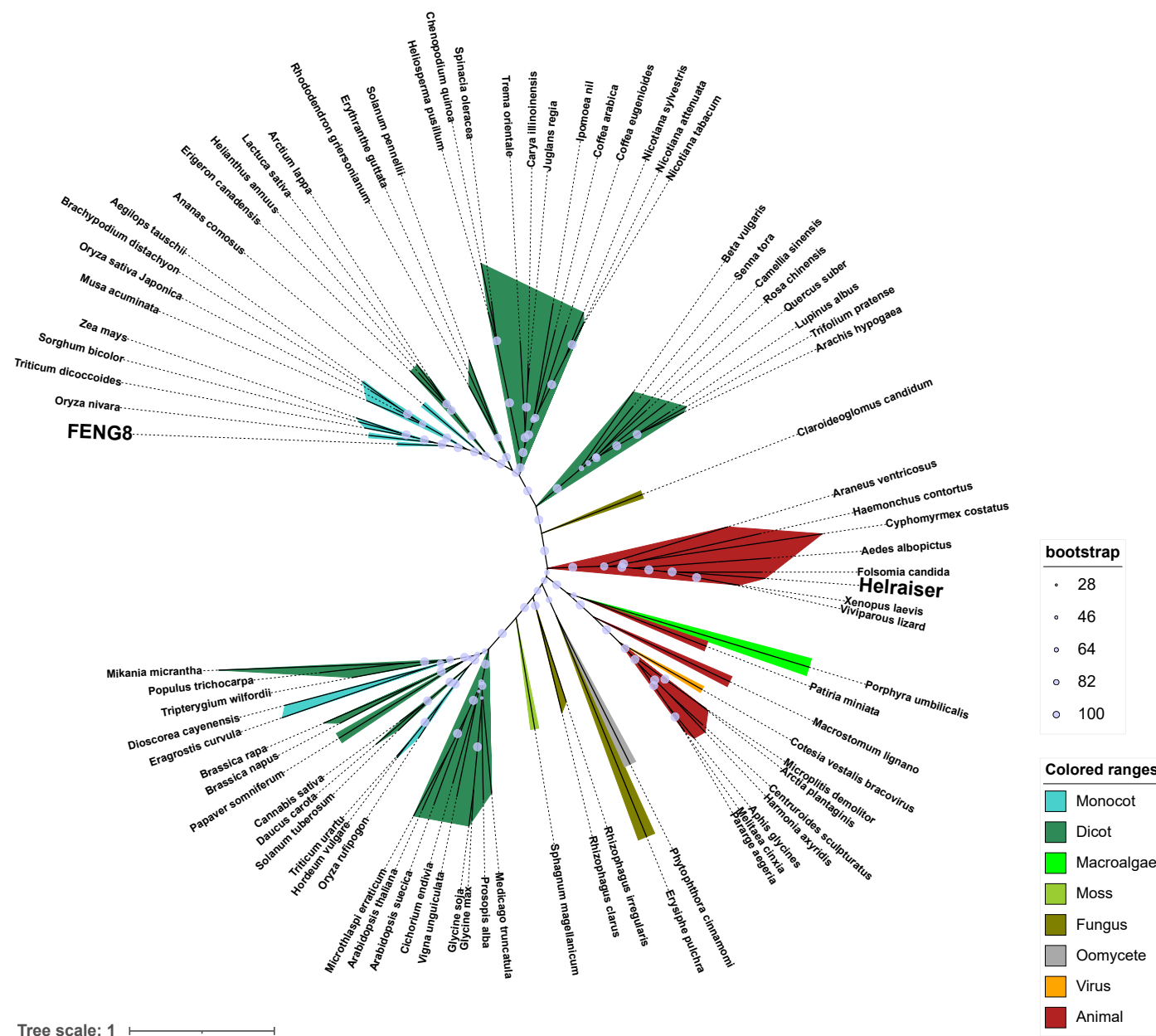
a**b****c****d****e**

Extended Data Fig. 6 | See next page for caption.

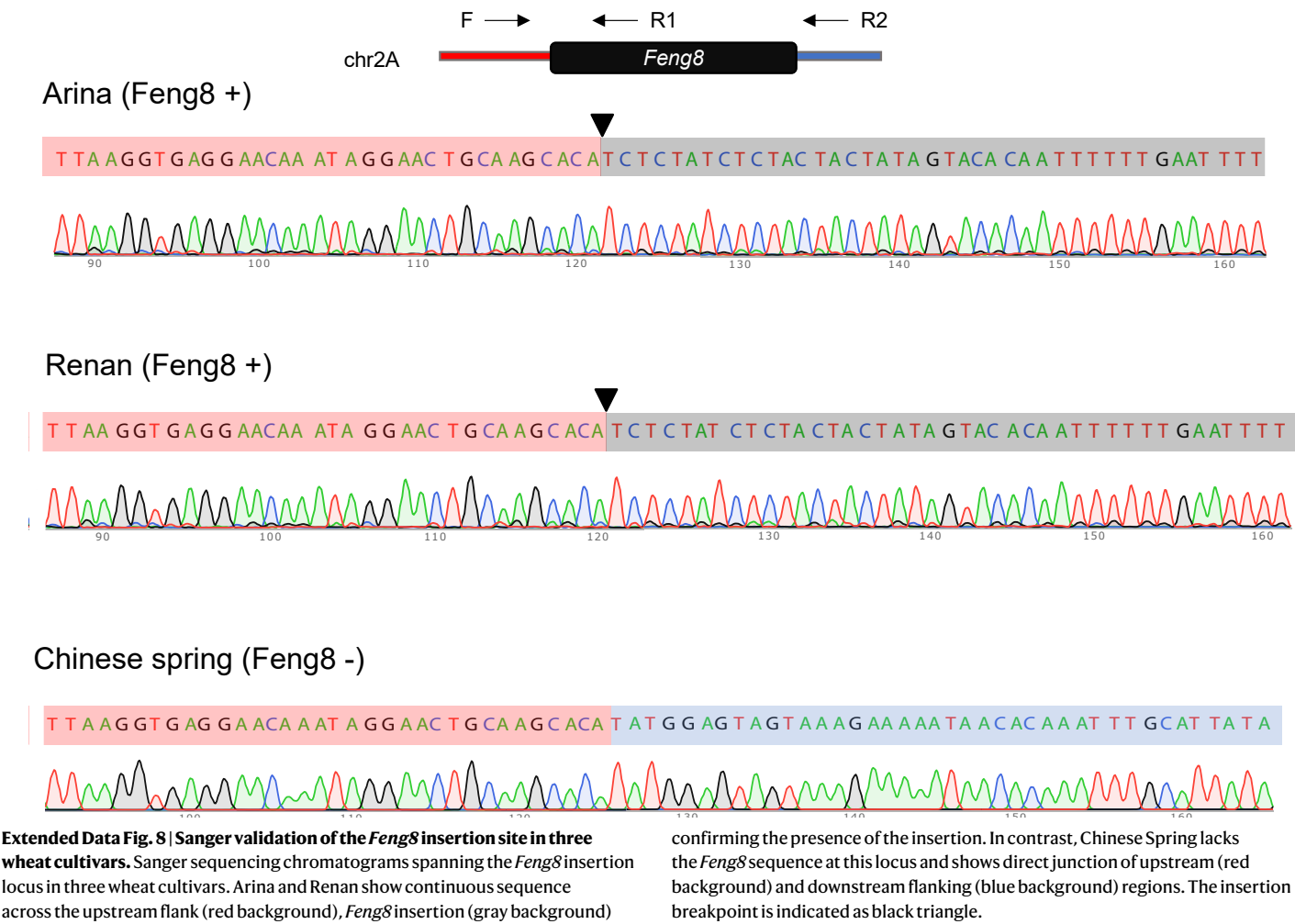
Extended Data Fig. 6 | Structural comparison of FENG8 and Helraiser.

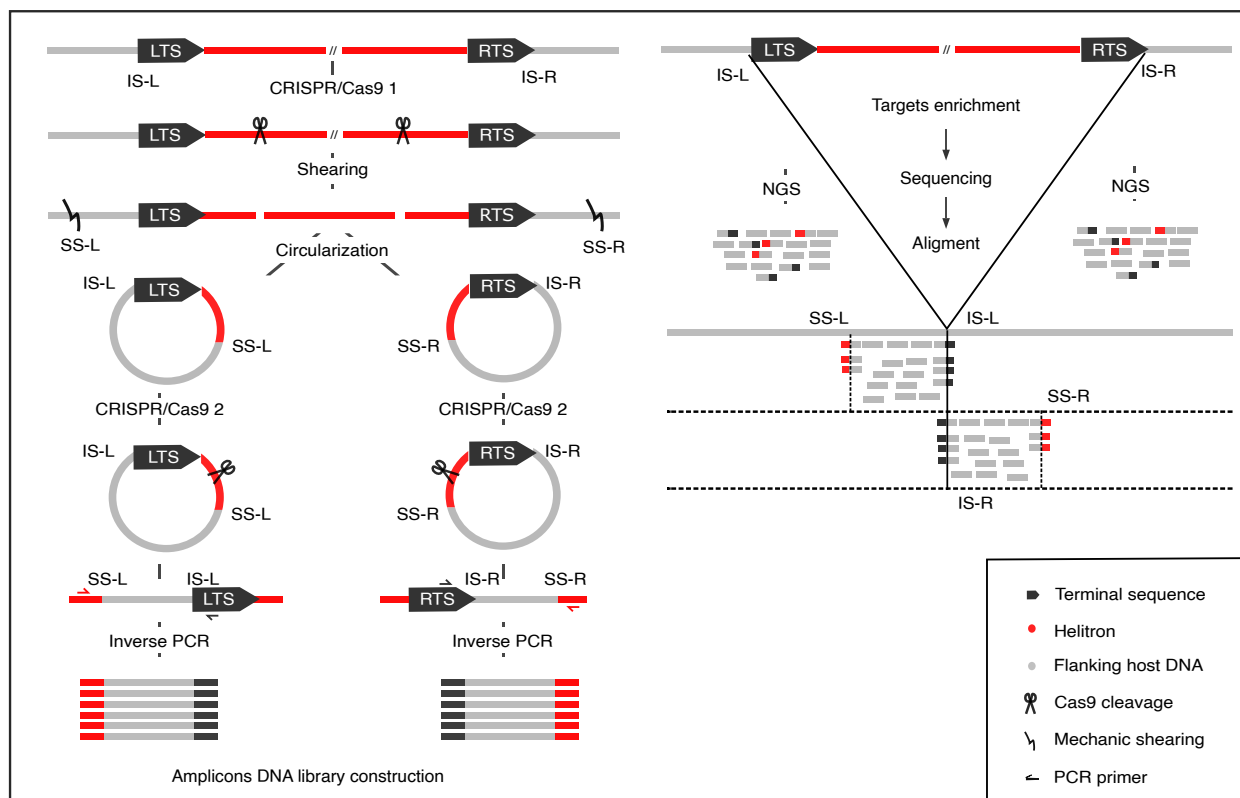
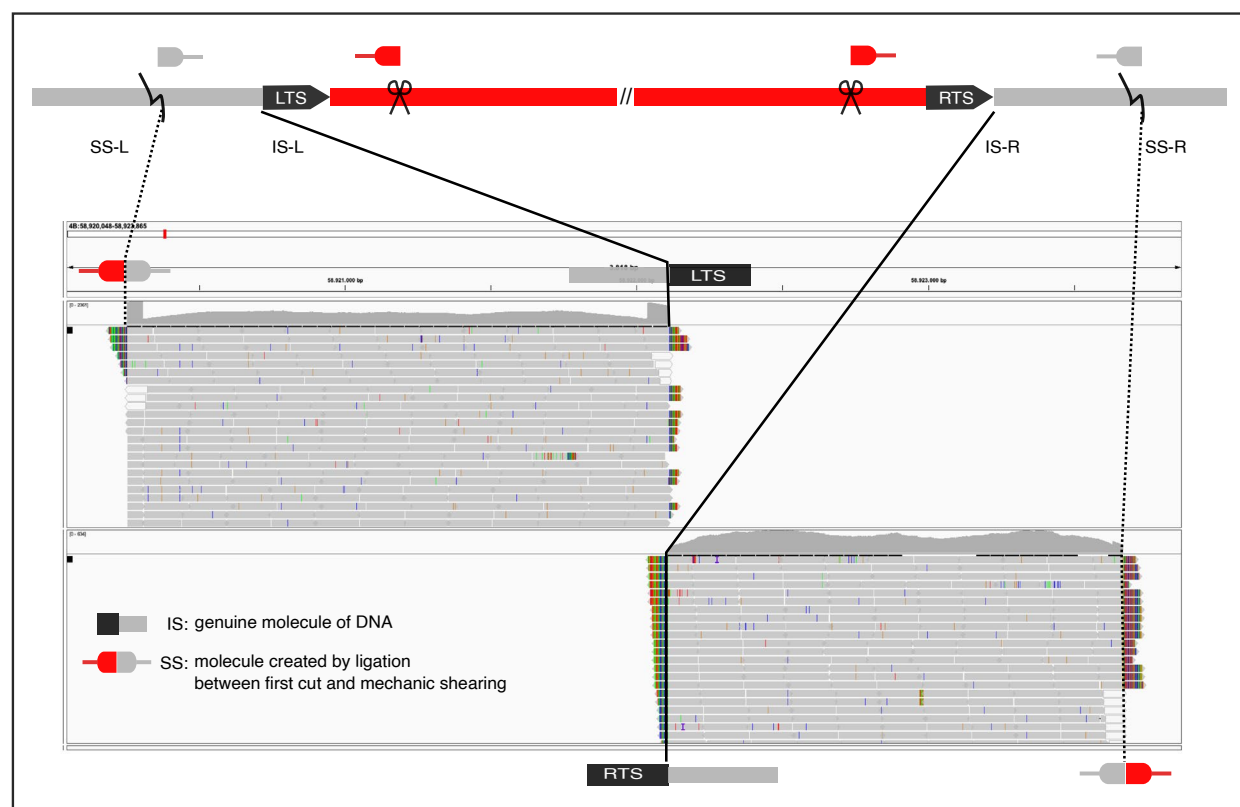
a, AlphaFold-predicted structure of FENG8, with colors indicating pLDDT confidence scores (upper panel). **b**, N-terminal domain of FENG8 (pink), with a disordered region of 264 amino acids highlighted in blue. **c**, Structural comparison

of the N-terminal domain in FENG8 (pink) and Helraiser (cyan). **d**, Structural comparison of the HUH domain in FENG8 (pink) and Helraiser (cyan). **e**, Structural comparison of the helicase domain in FENG8 (pink) and Helraiser (cyan).



Extended Data Fig. 7 | Phylogeny of FENG8 and other Helitron-like transposases. Maximum-likelihood phylogeny of the FENG8 transposase with Helitron-like transposases from 84 species. Detailed information for each protein can be found in the Supplementary Table 2.



a**b**

Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Identification of *de novo* Helitron insertions using PCIP-seq. a, Graphical representation of the PCIP-seq procedures. Genomic DNA is specifically cut with a CRISPR/Cas9 targeting the *Helitron* genomic sequence. The DNA is then mechanically sheared, generating random left and right shearing sites (SS-L and SS-R). The DNA is then end-repaired, circularized, and cut again by a second set of two CRISPR/Cas9 targeting *Helitron* sequences adjacent to *Helitron* LTS and RTS to specifically reopen circles encompassing left and right insertion sites (IS-L and IS-R) in two separated pools (left terminal

site [LTS] and right terminal site [RTS]). Fragments encompassing the IS and SS are then amplified by long-range inverse PCR using divergent *Helitron* targeting primer pairs. The resulting PCR products were then subjected to next generation sequencing (NGS) and the obtained sequence reads were mapped on the Chinese Spring IWGSC RefSeq⁵, revealing *Helitron* IS and their corresponding SS that allow the precise and sensitive identification of somatic *Helitron* insertions. **b**, Integrative Genome Viewer (IGV) screen capture of a sense *de novo* *Feng8* insertion captured in sample (#2883) in both LTS and RTS pools using PCIP-seq³⁸.

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	We used this open source software: Guppy v7.1.4 for basecalling
Data analysis	We used these software packages: minimap2 v2.26, nanomonsv v0.7, HELIANO, STAR v.2.7.10a, BAMscale, Trinity v.2.15.1, RSEM v.1.3.3, bowtie2, edgeR 7 v.4.0.3, R v.4.3.2, AlphaFold v2.3.1, PyMOL v2.1, MUSCLE v5.1, trimAl v1.4, IQ-TREE v2.12, BWA-MEM, SAMtools, Python code we wrote can be found here: https://github.com/Lijingtangbo

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 Xuan1 and Feng8 sequences have been submitted to NCBI GenBank with accession numbers PP376094 and PP376095. The sequencing data from this study have been submitted to the European Nucleotide Archive (ENA, www.ebi.ac.uk/ena/, accessed on ERP157467) under the project PRJEB72688. The whole genome

sequencing raw reads are under fastq accessions ERR12723706-ERR12723708, raw reads of eccDNA-Seq for Arina are under the accessions ERR12724408-ERR12724413. The raw reads of the RNA sequencing are under the accessions ERR12724677-ERR12724682. Positions of TaBW420K SNP markers on CS IWGSC RefSeq v2.1 are available under DOI: <https://doi.org/10.57745/YXNWZK>. The raw fastq files for the PCIP-seq experiments applied to the 24 RILs have been submitted to the European Nucleotide Archive (ENA, www.ebi.ac.uk/ena) under the projects PRJEB75199 and PRJEB108718

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	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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](https://www.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	For the experiments, at least 3 biological replicates were used (including transcriptomics).
Data exclusions	No data was excluded from the analyses
Replication	All experiments were carried out at least twice and with at least 3 biological replicates
Randomization	does not apply
Blinding	The eccDNA experiments on the RILs and the Xuan-Feng insertion analyses by PCIP were carried out by blinding. Meaning that the person performing the experiments did not know the genotype of the tested RILs. The IDs of the RILs were only revealed after the results were obtained.

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

Plants

Seed stocks

All wheat varieties were obtained from the Agroscope National Gene Bank

Novel plant genotypes

Arabidopsis plants were transformed using Agrobacterium and binary vectors. From each constructs two independent transformant lines were tested in our experiments.

Authentication

Plants of the Agroscope National Gene Bank are genotyped. All RILs used were genotyped.