Mobilization of Pack-CACTA transposons in Arabidopsis reveals the mechanism of gene shuffling

Marco Catoni[12]*, Thomas Jonesman[1], Elisa Cerruti[1] and Jerzy Paszkowski[1]*

Affiliations

[1] The Sainsbury Laboratory, University of Cambridge, Cambridge, United Kingdom

[2] School of Biosciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom

*Correspondance: jerzy.paszkowski@slcu.cam.ac.uk; marco.catoni@slcu.cam.ac.uk

Contact Information

Marco Catoni: marco.catoni@slcu.cam.ac.uk

Thomas Jonesman: thomas.jonesman@slcu.cam.ac.uk

Elisa Cerruti: elisa.cerruti@slcu.cam.ac.uk

Jerzy Paszkowski: jerzy.paszkowski@slcu.cam.ac.uk

Running title

Mobilization of Pack-TYPE transposons

Keywords:

Pack-CACTA, gene shuffling, genome stability, DNA methylation, epiRILs, transduplication

Abstract

Pack-TYPE transposons are a unique class of potentially mobile non-autonomous elements that can capture, merge and relocate fragments of chromosomal DNA. It has been postulated that their activity accelerates the evolution of host genes. However, this important presumption is based only on the sequences of currently inactive Pack-TYPE transposons and the acquisition of chromosomal DNA has not been recorded in real time. We have now for the first time witnessed the mobilization of novel Pack-TYPE elements related to the CACTA transposon family over several plant generations. Remarkably, these elements tend to insert into genes as closely spaced direct repeats and they frequently undergo incomplete excisions, resulting in the deletion of one of the end sequences. These properties constitute a mechanism of efficient acquisition of genic DNA residing between neighbouring Pack-TYPE transposons and its subsequent mobilization. Our work documents crucial steps in the formation *in vivo* of novel Pack-TYPE transposons and thus the mechanism of gene shuffling mediated by this type of mobile element.

Introduction

Autonomous transposable elements (TEs) encode all factors required for transposition, whereas non-autonomous elements react to factors provided *in trans* (Lisch, 2013). In DNA transposons, terminal inverted-repeat sequences (TIRs) of transposition-competent non-autonomous TEs are recognized by a transposase of a related autonomous element (Skipper et al., 2013). The extremities of Pack-TYPE elements include TIRs, while their internal sequences are derived from host chromosomes (Bennetzen, 2005; Jiang et al., 2011; Kawasaki and Nitasaka, 2004; Yu et al., 2000; Zabala and Vodkin, 2005). For example, Pack-MULEs are Pack-TYPE elements that populate the rice genome (Jiang et al., 2004), with approximately 3000 Pack-MULEs carrying segments of about 1000 different genes (Jiang et al., 2011). This heterogeneous population of transposons has had a major impact on the current organization of rice chromosomes and the evolution of rice genes (Jiang et al., 2004). Structures resembling Pack-TYPE transposons are also present in Arabidopsis (Gilly et al., 2014; Yu et al., 2000), where epigenetic suppression mediated by DNA methylation appears to render them immobile (Lister et al., 2008). Elimination of DNA METHYLTRASFERASE1 (MET1) activates the genome-wide transcription of TEs by erasing CpG methylation (Zilberman et al., 2007). Importantly, the absence of CpG methylation persists over many plant generations, even after restoration of MET1 activity (Catoni et al., 2017; Kankel et al., 2003; Saze et al., 2003). This observation prompted the construction of epigenetic recombinant inbred lines (epiRILs) from a cross between two isogenic parents: the wild type and a *met1* mutant deficient in CpG methylation (Reinders et al., 2009). F2 progeny with MET1 activity were further propagated by single-seed descent for eight generations, resulting in 68 epiRILs with mosaic methylation patterns (Reinders et al., 2009). Here, we screened for new TE mobilized in epiRILs, and we identified a new family of Pack-TYPE mobile elements with CACTA1-derived Terminal Inverted Repeats (TIRs). The study of real time mobilization of one of these new Pack-CACTA elements allowed proposing the mechanism of gene shuffling by this type of transposon.

Results

DNAseq of PCR-free genomic libraries for the 68 epiRILs was used to search for TE mobilization, which is usually associated with increased copy number (Chuong et al., 2017; Lisch, 2013). We looked for TE-annotated loci for which read coverage was at least 1.8-fold higher than in wild-type Col-0. We then compared putative activity levels of these TEs by the average increase in sequencing coverage. According to this analysis, retrotransposon EVADE (EVD) (Mirouze et al., 2009) was the most active, with an average 11.8-fold coverage increase (Supplementary Table 2). Surprisingly, the second most active element, with a 6.2-fold increase, had an unusually complex structure of two short terminal DNA stretches of 451 bp (AT4TE18505) and 337 bp (AT4TE18510) (annotated as a member of the ATENSPM3 family of DNA transposons) separated by a sequence annotated as gene AT4G07526 of unknown function (Figure 1A). The putative product of this gene showed no similarity to any known transposase and the overall transposon structure, residing on chromosome 4, resembled a non-autonomous Pack-TYPE element.

AT4TE18505 and AT4TE18510 contain TIRs beginning with 8 base pairs (CACTACAA) that are also a feature of the CACTA1 autonomous transposon (Miura et al., 2001), which is active in the epiRILs (Reinders et al., 2009) (Supplementary Table 2). A blast search with 150 bp of both terminal sequences identified nine additional elements in the Col-0 genome with similar structures and CACTA1-like termini (Figure 1B). These included conserved terminal sequences of 26 bp (Figure 1C) and a 3-bp target site duplication (TSD) (Supplementary Table 3). Their 150-bp terminal sequence identities ranged from 62 to 95.3% (Supplementary Table 4). Five of the elements shared related DNA sequences located between the CACTA1-like termini, while the remaining five elements carried sequences of different chromosomal origin (Figure 1B and D). Interestingly, not all of the captured sequences displayed similarities to Arabidopsis genes (Figure 1D) and DNA sequence databases did not reveal their origins. The TIR sub-terminal sequences of CACTA family transposons differ slightly between the 5’ and the 3’ ends (Frey et al., 1990). These differences were also present in the possibly active Pack-TYPE elements (Figure 1E) we discovered. Therefore, we called this family Pack-

52 CACTA, which in the Arabidopsis Col-0 accession consists of 10 TEs divided up, according to their
53 sequence similarities, into four groups (Figure 1B).

54 We found 50 new insertions of Pack-CACTA1a and 3 new insertions of Pack-CACTA2a in 8 and 3
55 epiRILs, respectively (Supplementary Table 5). In line epi26, Pack-CACTA1a and Pack-CACTA2a
56 appeared mobile (Supplementary Table 5). New transposed copies were found mostly within
57 euchromatic chromosomal regions (65%), resembling the insertion preference of CACTA1
58 (Supplementary Figure 2A). We validated new insertions of Pack-CACTA1a by locus-specific PCR,
59 choosing 11 random sites in epi26 and epi46 and two sites with Pack-CACTA2a insertions in epi14
60 and epi26 (Supplementary Figure 2B, C and Supplementary Table 6). Thus, both Pack-CACTA1a and
61 Pack-CACTA2a were currently transposing and generating TSD of three nucleotides (Supplementary
62 Data 1).

63 To better examine the timing and consequences of Pack-CACTA1a transposition, we characterized in
64 more detail five new insertion sites in 10 sibling plants each of epi26 and epi46 (Supplementary
65 Figure 3 A,B, C). We recorded the presence or the absence of Pack-CACTA1a in individual sibling
66 plants (Supplementary Figure 3D, E). The results are consistent with genetic segregation due to the
67 initial hemizygosity of recent insertions and/or the frequent excision of newly inserted transposons
68 (Figure 2A). In the case of transposon excision, we expected footprints that would be absent in the
69 corresponding wild-type locus. In different epi26 sibling plants, we rescued sequences of both intact
70 target loci and those containing footprints of the transposon (Figure 2B and Supplementary Data 2).
71 The results are consistent with high transpositional activity of Pack-CACTA1a in epiRILs during
72 inbreeding as well as somatic excision in tissues of individual plants.

73 Unexpectedly, in some plants of epi26 locus 4 (epi26_4) (Supplementary Figure 3F) we found a 264-
74 bp fragment of the CACTA1a 5' sequence flanked by 3 bp of TSD (Figure 2C and Supplementary Data
75 3). Ten epi26 plants with a Pack-CACTA1a insertion at this locus contained either the entire Pack-
76 CACTA1a or its deletion derivative (Supplementary Figure 3F). Therefore, it is most probable that the

77 deletion resulted from aberrant excision of a full copy from this location. Importantly, we did not
78 detect somatic “reversions” to a wild type-like locus in plants homozygous for the deleted version of
79 Pack-CACTA1a. The fragmented element was apparently immobilized at this new location (Figure
80 2C).

81 To investigate the frequency and features of aberrant excisions of Pack-CACTA1a, we examined a
82 further 64 sibling plants each of epiRILs epi26 and epi46, searching for additional deletion events at
83 five loci at which complete insertions of Pack-CACTA1a were previously found. To this end, we used
84 locus-specific primers that flanked each of the putative Pack-CACTA1a target sites (Supplementary
85 Figure 3C). Six independent deletion events were detected in 640 PCR reactions, two in epi26 and
86 four in epi46 (Figures 2D, Supplementary Figure 4). Therefore, generation of deletion derivatives of
87 this transposon appear to be quite common: in approximately 5% of plants or in 1% of loci
88 previously targeted by Pack-CACTA1a, and at least some of these are transgenerationally
89 transmitted (Figure 2D and Supplementary Figure 4C)

90 In the case of transposon insertion in an active gene, TIR deletion could stably alter expression of the
91 gene and/or alter the structure of its transcript. Indeed, in one case we detected a novel hybrid
92 mRNA initiated within a promoter residing in the remaining part of Pack-CACTA1a, the second exon
93 of the Pack-CACTA1a “passenger gene” was fused with the exon of a downstream chromosomal
94 gene (Figure 3 left panels and Supplementary Data 3). In a second case, the 3’ part of the transposon
95 was inserted in a gene exon; thus, its sequence formed part of the mature gene transcript (Figure 3
96 right panels and Supplementary Data 4).

97 The sequences of remaining fragments of the transposon showed that deletions occurred at
98 approximately equal frequencies at the 3’ (3 deletions) and 5’ terminals (4 deletions) (Figure 2C,D,
99 Supplementary Figure 4, Supplementary Data 5). This observation suggested that incorporation of
100 chromosomal DNA into a novel Pack-CACTA element requires the presence of two neighbouring
101 transposons inserted in the direct orientation. If such a pair of Pack-Type transposons, separated by

102 a short stretch of chromosomal DNA, undergoes complementary deletions of each of the elements,
103 a “hybrid” Pack-CACTA encompassing chromosomal DNA within the respective 5’ and 3’ terminals of
104 the former transposons would be generated (Figure 4A). To examine whether Pack-CACTA1a
105 integrates frequently at closely linked chromosomal locations, we designed a PCR-based screen to
106 recover chromosomal DNA stretches residing between putative neighbouring new inserts of Pack-
107 CACTA1a (Figure 4B). Obviously, this approach can only detect insertions separated by relatively
108 short stretches of chromosomal DNA. Therefore, it was surprising that numerous amplifications of
109 chromosomal DNA fragments were observed in the 128 DNA samples of epi26 and epi46 used
110 previously for the Pack-CACTA1a deletion screen (Figure 4C and Supplementary Figure 5). The
111 sequences of 28 different PCR products from 20 plants (Supplementary Figure 6 and Supplementary
112 Data 6, 7) all contained fragments of Arabidopsis genomic DNA of various sizes (38 to 2,137 bp)
113 between two Pack-CACTA1a copies. Remarkably, 27 PCR products revealed insertions of Pack-
114 CACTA1a as tandems in direct orientation and only one inversion (Supplementary figure 6). The
115 frequent formation by newly inserted copies of Pack-CACTA1a of direct repeats interspaced by short
116 stretches of chromosomal DNA, combined with the observed formation of terminal deletion
117 derivatives of the transposon, suggest that pairs of Pack-CACTA1a may form new “hybrid” elements.
118 Indeed, we recovered three deletion derivatives of Pack-CACTA1a among the 27 direct transposon
119 repeats (Supplementary Figure 6). Importantly, most neighbouring Pack-CACTA1a insertions were
120 separated by genic sequences (21 out of 28, Supplementary Figure 6); in one case three Pack-
121 CACTA1a insertions were located in one gene (Figure 4D).

122 Obvious extrapolations from the observed properties of active Pack-CACTA1a provide crucial clues
123 as to the mechanism by which Pack-TYPE transposons acquire and relocate stretches of
124 chromosomal DNA (Figure 4A). However, detection of a newly formed, transpositionally active Pack-
125 CACTA1a derivative would require the screening of several thousand plants harbouring currently
126 active parental elements.

127 As an alternative, we searched for structural variants of Pack-CACTA1a elements in the newly
128 assembled genome of the Arabidopsis Ler-0 accession and retrieved eight analogous elements
129 (Supplementary Figure 7 and Supplementary Table 7). Of these, the sequences of three full-length
130 elements and two deletion derivatives were closely related to Pack-CACTA1a of Col-0, with sequence
131 identities of 73 to 90% (Supplementary Table 7). The sequences of the remaining three copies of Ler-
132 0 Pack-CACTA elements resembled Pack-CACTA1a only in the first 992 bp and the last 325 bp with
133 76% and 72% identities, respectively. Remarkably, the internal 1,181 bp showed no similarity to the
134 "parental" element but contained a 132-bp stretch with 73% identity to the second exon of gene
135 AT3G48620 encoding a putative protein of the outer membrane family (OMP85) (Figure 4E). This
136 novel structure of a Ler-0 relative of Pack-CACTA1a suggests that a Pack-CACTA transposon in this
137 Arabidopsis accession could capture and move chromosomal DNA by a mechanism compatible with
138 that we recorded in real time for Pack-CACTA1a transposition in Col-0.

139 In rice, previous analyses of Pack-MULE transposons considered only complete elements with both
140 terminal repeats encompassing various fragments of rice chromosomal DNA (Jiang et al., 2004). We
141 reanalysed available rice genomic sequences for intermediates of chromosomal DNA acquisition by
142 rice Pack-MULEs, such as transposon deletions due to aberrant excision events. We applied a blast
143 search across the rice genome using as a query 2,853 previously retrieved pack-MULE intact
144 elements (Jiang et al., 2004) and filtered results to recover incomplete elements with only one
145 terminal sequence (see Methods). Of the 2,151 entries meeting these criteria (Supplementary Data
146 8), 269 (13 %) Pack-MULE deletion derivatives resided less than 10 kb from an intact Pack-MULE
147 element (Figure 4F). Pack-MULEs consisting of terminal sequences flanked by different target site
148 duplications, making up 17% of the Pack-MULE population, were previously discarded as false
149 positives (Jiang et al., 2004). However, considering the deduced mechanism of DNA acquisition by
150 Pack-TYPE elements, part of these putative transposons may represent cases of newly generated but
151 not yet relocated hybrid Pack-MULEs consisting of terminal sequences from two different parental
152 elements that underwent aberrant excisions (Figure 4G). Thus, the abundant Pack-MULEs of the rice

genome, which have contributed significantly to its current organization, seem to include numerous examples of putative transposon intermediates consistent with the scheme of chromosomal DNA acquisition during real-time Pack-CACTA mobilization in Arabidopsis (Figure 4A).

Discussion

A massive contribution of Pack-TYPE transposon activity to the evolution of plant genes and genomes was deduced from analyses of rice, Brassica and soybean (Alix et al., 2008; Hanada et al., 2009; Jiang et al., 2004; Jiang et al., 2011; Zabala and Vodkin, 2005). It was postulated that repair of nicks and gaps created at the transposon insertions sites or arising due to the element structures might be responsible for the acquisition of chromosomal DNA (Bennetzen and Springer, 1994; Engels et al., 1990; Yamashita et al., 1999). Such DNA-repair mechanisms would operate prior transposon relocation and the newly generated elements would than move to different chromosomal locations only after incorporation of chromosomal DNA. Here, when studying transposition properties of Pack-CACTA1a elements *in vivo*, we found that acquisition of chromosomal DNA seems to be tightly linked to the transposition. Pack-CACTA1a transposition has a marked tendency to produce closely spaced repeats in direct orientation. Subsequently, complementary aberrant excisions of the elements within these repeats, leaving behind one of the two TIRs, generate a novel “hybrid” Pack-CACTA that incorporates the chromosomal DNA that separated the two neighbouring insertions (Figure 4A).

This quite simple and efficient mechanism of chromosomal DNA acquisition relies on the combination of known characteristics of particular DNA transposons. For example, maize Ac/Ds, P-element of *Drosophila,* Tc1 transposon in *Caenorhabditis* and *Sleeping Beauty* in mouse tend to transpose to nearby locations, termed “local hopping” (Bancroft and Dean, 1993; Carlson et al., 2003; Fischer et al., 2003; Skipper et al., 2013; Zhang and Spradling, 1993). However, the orientation of elements in such integration arrays are random and not preferentially direct, as observed for Pack-CACTA1a. Incomplete or aberrant excision is known for CACTA-like transposable elements in soybean (Xu et al., 2010), while maize chromosomal aberrations, like breakages and fusions, were

178 previously associated with the transposition of TIRs belonging to two different neighbouring Ac/Ds
179 elements (Ralston et al., 1989). It is interesting that Pack-CACTA1a combines both properties, i.e.
180 efficient formation of closely linked tandem insertions in direct orientation and frequent aberrant
181 excisions that lead to complementary deletion of one of the two TIRs of each pair. The remaining
182 terminals of the neighbouring transposons encompass a new, single element incorporating the
183 intervening chromosomal DNA.

184 The incorporation of gene fragments and their duplication by transposition to new genomic
185 positions has been suggested for certain TE families, including MULEs, CACTAs and HELITRONs (Lisch,
186 2013). This "transduplication" process seems to be very frequent in rice and can be clearly assigned
187 to the historical activity of Pack-MULEs (Juretic et al., 2005). Importantly, transduplicated genes
188 seem to be under purifying selection (Hanada et al., 2009) or serve regulatory functions (Juretic et
189 al., 2005). This implies that Pack-TYPE transposon mobilization may have a direct and profound
190 influence on gene evolution, generating new gene functions or regulatory activities by shuffling parts
191 of various coding regions across the genome. Taking into account the high rates of nearby
192 directional insertion and aberrant excision observed here for the Pack-CACTA1a element, the "TE-
193 assisted" origin of genes may be strongly underestimated, especially in plants with genomes larger
194 than Arabidopsis.

195 In hypomethylated Arabidopsis inbred lines, we followed for several plant generations mobilization
196 of novel Pack-TYPE elements related to the family of CACTA transposons and we witnessed the initial
197 steps of chromosomal DNA acquisition. Our unique data *in vivo* allowed proposing the mechanism of
198 gene shuffling by this type of transposon. This mechanism is directly linked to their transposition
199 properties and therefore not only depending by DNA repair activities of the host, as it has been
200 postulated so far.

201

202 # Methods

203 **Plant material**

204 The *met1-3* derived epiRIL population was described previously (Reinders et al., 2009). From this
205 population, the 9th inbred generation was grown to produce the 68 epiRILs used in this study. Col-0
206 and 2nd generation *met1-3* plants (Saze et al., 2003) were used as controls. Unless indicated
207 otherwise, plants were grown in soil under long-day conditions (21°C, 16 h light, 8 h dark).

208 **DNA extraction and library preparation**

209 Arabidopsis seeds were sown on ½ MS (Plant) medium plates and grown for 14 days before use.
210 Approximately 10 seedlings per epiRIL were harvested (approx. 100 mg fresh weight) and DNA
211 extracted using the Qiagen DNeasy Plant Mini Kit following the manufacturer's instructions (Qiagen
212 N.V., Hilden, Germany). Before library preparation, DNA was fragmented to an average size of 350
213 bp using 24 cycles of 30 s with a Bioruptor Diagenode sonication device. The genomic DNA
214 sequencing libraries were prepared using the TruSeq DNA PCR-Free LT Library Prep Kit following the
215 manufacturer's instructions (Illumina, San Diego, Calif.), starting from 1.1 µg of fragmented DNA.
216 Libraries were validated using High Sensitivity D1000 screentape on a 2200 Tapestation instrument
217 (Agilent technologies, Santa Clara, CA) and a LightCycler 480 Instrument II using the LightCycler 480
218 SYBR Green I Master mix (Roche, Basel, Switzerland).

219 **DNA sequencing, reads mapping and peak call**

220 The DNA libraries were sequenced with 2 x 76-bp paired-end reads with a minimum of 10x coverage
221 (24x on average) on an Illumina NextSeq 500 using the High-Output Flow Cell configuration. The raw
222 reads were trimmed using Trimmomatic to remove adapter sequences. Reads with an averaged
223 value of at least 15 in a 4-nt window were trimmed from both ends. After trimming, reads pairs with
224 at least one mate shorter that 36 bp were discarded. The remaining sequences (on average 95% of
225 raw reads) were aligned with bowtie2 (Langmead and Salzberg, 2012) with the –no mixed and –non

226 deterministic option, against the reference TAIR10 Arabidopsis genome (www.arabidopsis.org).
227 Metrics of the sequencing analysis are displayed in Supplementary Table 8. Genome coverage was
228 calculated for each genome position using genomecov (bedtools v2.26.0) and normalized to the
229 Arabidopsis genome and library size. Arabidopsis genomic regions with a coverage of more than 2.5-
230 fold or less than 0.2-fold of the average genome coverage in the Col-0 wild-type control were
231 masked to avoid calling peaks in regions with high copy number or insufficient coverage (e.g. plastids
232 DNA, ribosomal repeats, microsatellites). Peak call was done for each epiRIL mapped bam file in
233 comparison to the Col-0 control condition with macs2 (https://github.com/taoliu/MACS/), setting
234 the fragment size to 75 and extended size to 400 bp, with the options –B, -SPMR and –no model and
235 Columbia-0 as control. The Log based background subtraction was performed with macs2 bdgcmp
236 with "logLR" as model and a *P*-value threshold of 0.00001.
237 The peak lists obtained from all epiRILs were merged and peaks closer than 5 kb were joined
238 together in a gff file using R (https://www.R-project.org/). Then, HTseq (Anders et al., 2015) was
239 used to count reads at each peak for all conditions analysed and the raw count was normalized to
240 peak length and library size to obtain FPKM values. Fold change (FC) of DNA coverage was calculated
241 for each peak by the ratio between FPKM values in the tested condition and the Col-0 wild-type
242 control. Peaks overlapping TEs (TAIR10 annotation) were further filtered to contain only regions with
243 more than 1.8-fold change difference in at least one epiRIL.
244 **Detection of TE insertions**
245 New putative TE integrations were detected in each epiRIL with the Transposon Insertion Finder (TIF)
246 algorithm (Nakagome et al., 2014), using the first and last 17 bp of each TE tested and the
247 Arabidopsis TAIR10 genome as reference. For each sample, the paired fastq files obtained from the
248 genome sequencing were concatenated in a single file and used as input for TIF. The new putative
249 insertion loci obtained from the analysis were filtered to exclude false positive calls in Col-0
250 sequenced samples.
251 **PCR and cloning**

252 To confirm new transposon insertions, PCRs were carried out using a transposon-specific primer and
253 a primer flanking the new insertion. Amplification of the entire locus with new Pack-CACTA1a
254 integration was performed using two primers designed on the sequences flanking the new insertion.
255 All PCR reactions were carried out using GoTaq enzyme (Promega, Madison, WI, USA), with
256 extension time adjusted to the expected size of the fragment amplified, following the recommended
257 manufacturer's instructions. All PCR products were extracted from the gel using a Qiagen Gel
258 Extraction Kit and eluted in 30 µl of water. The purified PCR products were ligated into a pGEM-T
259 plasmid (Promega) following the manufacturer's instructions and 2.5 µl of ligated sample used to
260 transform E.coli DH5α cells (50 µl). Colony PCRs were performed starting from overnight grown *E.*
261 *coli* colonies using the universal M13 forward and reverse primers. The amplified products were
262 SANGER sequenced by Sigma-Aldrich (Merck, Darmstadt, Germany). Primer sequences are listed in
263 Supplementary Table 9.

264 **RT PCR**

265 Total RNA was extracted from 150 mg of fresh leaf tissue or green siliques using the Trizol
266 (Invitrogen) method according to the manufacturer's instructions and resuspended in 50 µl of water.
267 Total RNA (10 µl) was treated with RQ1 DNase (Promega) and reverse transcribed using Superscript
268 II (Invitrogen) following the manufacturer's instructions. RT-PCR was carried out using specific
269 primers designed on exons of the Pack-CACTA1a transposon and the gene target of integration
270 (Supplementary Table 9).

271 **BLAST search**

272 BLAST analyses were carried out using the NCBI web interface
273 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or the locally installed BLAST+ v2.2.25 (Camacho et al.,
274 2009), depending on the analysis. When the local version was used, blastn was run with the options
275 –evalue 1e-5 and -max_target_seqs 100 where not specified otherwise.

276 **Pack-CACTA analysis**

277 Additional Pack-CACTA elements were detected in the Arabidopsis genome by blasting the 26-bp
278 core TIR sequences (Figure 1C) of the active Pack-CACTA1a element to the Arabidopsis reference
279 genome (TAIR10). The blastn (BLAST+ v2.2.25) was run with the options "-task blastn-short" and "-
280 word_size 7". The retrieved 18 matching sequences were imported in Geneious v7.1.7 (Biomatters)
281 and manually checked to remove sequence derived from autonomous CACTA elements.

282 In order to identify Pack-CACTA elements in the genome of Ler-0 accession, the entire Pack-CACTA1a
283 sequence was blasted on the PacBio Ler-0 assembly (http://www.pacb.com/). The sequences of the
284 matching hits were retrieved from the assembly and aligned with Geneious v7.1.7 (Biomatters),
285 using the MUSCLE algorithm and default parameters. Consensus trees were obtained from the
286 alignment using the Temura-Nei genetic distance and Bootstrap resampling of 1000 replications.
287 TIRs and TSDs were manually annotated.

288 **Rice Pack-MULE analysis**

289 The annotation of rice Pack-MULEs was obtained by Jiang et al. (Jiang et al., 2011) and the sequence
290 of each element was retrieved from the rice genome (http://rice.plantbiology.msu.edu/, release 6.0)
291 using R. The sequences obtained were blasted against the same reference genome with blastn
292 (BLAST+ v2.2.25) and the options–evalue 1e-5 and -outfmt 6. The list of 1,497,458 blast hits was
293 filtered by alignment length > 500 and evalue < 1e-5, selecting only blast hits with "query_start < 20"
294 or "query_end > (query_length – 20)" in order to select only results including at least one of the TIR
295 sequences. The list was further filtered by excluding any sequence overlapping the original set of
296 Pack-MULEs previously described (Jiang et al 2011). Remaining overlapping sequences were merged
297 to generate the final list of 2,151 partial pack-MULE elements (Supplementary Data 8).

298 **Availability of data and materials**

Sequencing data have been deposited in Gene Expression Omnibus under the accession number (sequencing data are under submission process). All the other data generated or analysed during this study are included in this published article or available under request.

Acknowledgments

Funding

This work was supported by the European Research Council (EVOBREED) [322621] and a Gatsby Fellowship [AT3273/GLE].

Disclosure Declaration

The authors declare that they have no competing interests

Author contributions

MC and JP designed the experiments; TJ prepared nucleic acid libraries; MC, TJ, and EC performed validation of new transposon insertions and their arrangement, MC performed all other experiments and analysis of the genomic data; MC and JP wrote the paper.

314 # Figure legends

315

316 **Figure 1. A novel non-autonomous pack-TYPE transposon related to CACTA becomes mobile in**
317 **epiRILs.**

318 A) Copy number changes in epiRILs of a Pack-TYPE element (Pack-CACTA1a) residing on chromosome
319 4. EpiRIL numbers are indicated left of each track. Differences to reference Col-0 control in log-
320 transformed coverage of reads plotted in the interval of -1 to 1 are mapped to the Pack-CACTA1a for
321 each track. TAIR10 annotation of genes (blue) and TEs (red) is displayed at the bottom.

322 B) Structure of Pack-CACTA family. Four groups of elements are marked in different colours with
323 exons represented by a darker hue. The phylogenetic tree was obtained using alignment of full-
324 length sequences. Numbers at each node indicate bootstrap support values of 1000 replications. The
325 5' terminal inverted repeat (5' TIR) is marked as a red triangle and 3' TIR as a green triangle. Target
326 site duplications of 3 bp (TSD) are marked as black arrows.

327 C) Sequence logo (http://weblogo.berkeley.edu/logo.cgi) obtained for alignment of TIRs of the 10
328 members of the Pack-CACTA family.

329 D) Chromosomal origin of Pack-CACTA sequences. The percentage identities to Arabidopsis genes
330 are indicated for each area.

331 E) Sequence relationships of the TIRs of Pack-CACTA groups. CACTA1 is annotated as AT2TE20205,
332 CACTA2 as AT1TE42210, CACTA3 as AT2TE18415, and CACTA4 as AT1TE36570. Numbers at each
333 node indicate bootstrap support values of 1000 replications.

334

335 **Figure 2. Hallmarks of Pack-CACTA1a excisions.**

336 A) Excisions of Pack-CACTA1a from locus 4 of epiRIL26 (epi26_4) in five epi26 plants. Gel-separated
337 PCR products were obtained with primers depicted in panel C as red or blue arrows. Lines marked as
338 "Pack-CACTA +" have a locus 4 containing a Pack-CACTA1a insert (indicated as "TE insertions") and in
339 plants marked as "Pack-CACTA –"Pack-CACTA1a underwent excision, resulting in "empty sites". The
340 wild-type Col-0 sample is marked as WT.

341 B) DNA sequences of PCR amplified "empty sites" displayed in panel A. The sequence of the initial
342 epi26_4 containing the TE insertion is provided in the first line and TSD of 3 bp is underlined. Letters
343 A-E left of the sequences correspond to the labelling in panel A.

344 C) Identification of an aberrant excision of Pack-CACTA1a from epi26_4. Two lines of the gel, marked
345 as 4Δ, represent epi26 plants that underwent aberrant excisions resulting in terminal deletion.

346 D) Identification of an additional aberrant excision event of Pack-CACTA1a (left panel) at a different
347 locus of epi26 plants (epi26_1) and its transgenerational inheritance (right panel). All symbols and
348 markings are as in panel C. The line marked 1Δ represents a plant homozygote for a transposon
349 insertion at epi26_1, in which only one Pack-CACTA1a was aberrantly excised. The remaining full-
350 length transposon is not visible due to PCR competition. The ladder is 1 kb NEB.

351

352

353 **Figure 3. Deletion derivatives of Pack-CACTA1a alter gene transcripts.**

354 A) Structures of new Pack-CACTA1a inserts that underwent aberrant excisions. Deleted parts of
355 Pack-CACTA1a are in grey. Exons are in darker and introns in lighter hues. Primers used for RtPCR are
356 shown as blue and red arrows. Deduced, novel mRNAs are displayed at the bottom, with green and
357 red triangles representing initiation and termination codons, respectively.

358 B) Results of RT-PCR performed on single epi26 homozygous (HM), heterozygous (HET) or wild type-
359 like (empty) plants for the corresponding Pack-CACTA1a deletion derivatives depicted in A. The
360 combination of primers used is shown on the left. Col-0 wild-type plants are shown as control (WT).
361 Amplification of ubiquitin mRNA (UBQ) and samples without reverse transcriptase (RT-) were used
362 as controls. Two biological replicates per genotype were used.

363

364 **Figure 4. High frequency of Pack- CACTA1a insertion at some loci support a model for DNA capture**
365 **in new pack-TYPE transposons.**

366 A) Possible mechanism of genic DNA capture by Pack-CACTA1a. New transposon insertions form a
367 closely spaced tandem array. Subsequently, neighbouring elements undergo aberrant excisions
368 resulting in complementary, terminal deletions, which form a novel hybrid Pack-CACTA element that
369 incorporates chromosomal DNA which separated the initial closely spaced insertions.

370 B) PCR strategy used for detection of neighbouring insertions of Pack-CACTA1a. Black arrows
371 indicate the primers.

372 C) Results of a PCR screen for neighbouring insertions of Pack-CACTA1a in 30 epi26 plants. Cloned
373 fragments corresponding to adjacent insertions are marked with black arrows. The full set for all
374 tested plants is displayed in Supplementary Figure 5.

375 D) Arrangement of Pack-CACTA1a copies in four representative loci containing tandem insertions
376 (data for other loci are presented in Supplementary Figure 6). The orientation of Pack-CACTA1a
377 copies is indicated with an arrow (5’ to 3’ direction). Red and yellow rectangles represent genes and
378 their transcripts, respectively.

379 E) Structure of a Col-0 Pack-CACTA1a-related element identified in the Ler accession. The percentage
380 sequence identities with Pack-CACTA1a and the acquired gene AT3G48620 are indicated.

381 F) Examples of rice Pack-MULE terminal deletion derivatives neighbouring intact Pack-MULE
382 elements. Black triangles represent TIRs. Blue and red arrows represent 9 bp TSD (different colours
383 correspond to different sequences). Red rectangles represent coding regions of genes.

384 G) Example of neighbouring Pack-MULE deletion derivatives compatible with the formation of a
385 novel hybrid Pack-MULE element. Symbols and colours as in F.

386

387 ## Supplementary Figure legends

388 **Supplementary Figure 1**

389 A) Copy number changes in epiRILs of EVADE' retrotransposon. EpiRIL numbers are indicated at the
390 left of each track. Log-transformed coverage difference of mapped reads at a locus between each of
391 the indicated epiRIL and the reference Col-0 control. In each track, the relative coverage is plotted in
392 intervals of 0 to 1. The track below indicates the TAIR10 annotation.

393 B) Copy number changes in epiRILs of DNA transposon CACTA. EpiRIL numbers indicated at the left
394 of each track. Log-transformed coverage difference of mapped reads at a CACTA1 locus between
395 each of the indicated epiRILs and the reference Col-0 control. In each track, the relative coverage is
396 plotted in intervals of -1 to 1. The decrease in coverage at some epiRILs observed is consistent with
397 excision events occurring for DNA transposons. The track below indicates the TAIR10 annotation.

398 C) Bar plot illustrating fold change (FC) in copy number of EVADE' (upper plot) and CACTA1 (lower
399 plot) transposons in each epiRIL line compared to Col-0. Blue bars represent DNA reads coverage
400 and red bars represent number of novel "split reads".

401

402 **Supplementary Figure 2**

403 A) Chromosomal distribution of new insertions of CACTA1 and Pack-CACTA elements in Arabidopsis
404 EpiRILs. Promoters were defined as stretches of maximum 1 kb DNA upstream of transcription starts
405 of genes.

406 B) Location of primers used for PCR validation of Pack-CACTA new insertions.

407 C) Validation of new Pack-CACTA insertions in epi14 (14) epi26 (26) and epi46 (46). DNA from the
408 original parental *met1-3* (M3) and Col-0 wild type (WT) are shown as negative controls.

409

410 **Supplementary Figure 3**

411 A) The positions of the loci amplified and cloned are indicated with numbers. The original Pack-
412 CACTA1a copy on chromosome 4 is marked by a blue line.

413 B) Distribution of new Pack-CACTA1a insertions in the epi46 genome (marked as in A).

414 C) Strategy for PCR amplification of loci with a new Pack-CACTA1a integration. Locus-specific primers
415 (red arrows) were designed for each new insertion studied.

416 D) Amplification of five Pack-CACTA1a insertions in epi26 single plants. The primer combination used
417 is represented by arrows on the right (for explanation see panel C). The black triangles mark the
418 sequenced fragments confirming the presence of novel insertions of full-length Pack-CACTA1a.

419 E) Amplification of five Pack-CACTA1a insertions in epi46 single plants (marking as in panel D).

420 F) Amplification of the locus 4 in additional epi26 plants with a Pack-CACTA1a insertion, showing the
421 presence of partially deleted transposons. A red triangular arrow indicates the sequenced Pack-
422 CACTA1a deleted fragment. Description as in D.

423 **Supplementary Figure 4**

424 Formation of Pack-CACTA1a deletion derivatives in loci epi46_1 (A) and epi46_5 (B). PCR
425 amplification products with deletion derivatives as marked by red triangles (left panels) and their
426 schematic representations according to the obtained sequences (pictures on the right). Primers used
427 are marked as red arrows. The ladder is 1 kb NEB (line M). Line WT represents Col-0.

428 C) Formation and transgenerational transmission of a Pack-CACTA1a deletion derivative in locus
429 epi26_2. The primers used are displayed as red and blue arrows. Red triangular arrows indicate the
430 sequenced Pack-CACTA1a deletion derivative. Other markings as in A and B.

431 **Supplementary Figure 5**

432 Results of PCR-based screening for closely spaced tandem insertions of CACTA1a (the strategy and
433 primers as in Figure 4B). (A) Gel-separated PCR products representing 64 plants of epi26. (B) PCR
434 products representing 64 plants of epi46. Red arrows point to samples subjected to sequencing. The
435 upper gel in A represents a repetition of samples displayed in Figure 4B (note the much shorter
436 separation of the PCR products in the gel presented here).

437 **Supplementary Figure 6**

438 Schematic maps of 28 DNA sequences obtained from the 20 PCR reactions marked in Supplementary
439 Figure 5. Red and green triangles symbolize 5' and 3' TIR, respectively. The positions of primers used
440 for amplification are shown as black arrows and explained in Figure 4B. Closely spaced pack-
441 CACTA1a insertions are predominantly arranged as direct repeats (27 examples of head-to-tail and
442 only 1 head-to-head, epi26.18). Arabidopsis genomic DNA is displayed in grey and its coordinates
443 orientation, and annotation are reported on the right.

444 **Supplementary Figure 7**

445 Structure of Pack-CACTA family in the Ler-0 accession of Arabidopsis. Exons are represented by
446 darker hue. The phylogenetic tree was obtained using the alignment of full-length sequences and
447 the numbers at each node indicate bootstrap support values of 1000 replications. Other symbols as
448 in Figure 1B.

449 ## References

450 Alix, K., Joets, J., Ryder, C.D., Moore, J., Barker, G.C., Bailey, J.P., King, G.J., and (Pat) Heslop-Harrison,
451 J.S. (2008). The CACTA transposon Bot1 played a major role in Brassica genome divergence and gene
452 proliferation. Plant J. *56*, 1030–1044.

453 Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-
454 throughput sequencing data. Bioinformatics *31*, 166–169.

455 Bancroft, I., and Dean, C. (1993). Transposition Pattern of the Maize Element Ds in Arabidopsis
456 Thaliana. Genetics *134*, 1221–1229.

457 Bennetzen, J.L. (2005). Transposable elements, gene creation and genome rearrangement in
458 flowering plants. Curr. Opin. Genet. Dev. *15*, 621–627.

459 Bennetzen, J.L., and Springer, P.S. (1994). The generation of Mutator transposable element
460 subfamilies in maize. TAG Theor. Appl. Genet. Theor. Angew. Genet. *87*, 657–667.

461 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L.
462 (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

463 Carlson, C.M., Dupuy, A.J., Fritz, S., Roberg-Perez, K.J., Fletcher, C.F., and Largaespada, D.A. (2003).
464 Transposon mutagenesis of the mouse germline. Genetics *165*, 243–256.

465 Catoni, M., Griffiths, J., Becker, C., Zabet, N.R., Bayon, C., Dapp, M., Lieberman-Lazarovich, M.,
466 Weigel, D., and Paszkowski, J. (2017). DNA sequence properties that predict susceptibility to
467 epiallelic switching. EMBO J. e201695602.

468 Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements:
469 from conflicts to benefits. Nat. Rev. Genet. *18*, 71.

470 Engels, W.R., Johnson-Schlitz, D.M., Eggleston, W.B., and Sved, J. (1990). High-frequency P element
471 loss in Drosophila is homolog dependent. Cell *62*, 515–525.

472 Ferguson, A.A., Zhao, D., and Jiang, N. (2013). Selective Acquisition and Retention of Genomic
473 Sequences by Pack-*Mutator*-Like Elements Based on Guanine-Cytosine Content and the Breadth of
474 Expression. Plant Physiol. *163*, 1419–1432.

475 Fischer, S.E.J., Wienholds, E., and Plasterk, R.H.A. (2003). Continuous exchange of sequence
476 information between dispersed Tc1 transposons in the Caenorhabditis elegans genome. Genetics
477 *164*, 127–134.

478 Frey, M., Reinecke, J., Grant, S., Saedler, H., and Gierl, A. (1990). Excision of the En/Spm
479 Transposable Element of Zea-Mays Requires 2 Element-Encoded Proteins. Embo J. *9*, 4037–4044.

480 Glusman, G., Severson, A., Dhankani, V., Robinson, M., Farrah, T., Mauldin, D.E., Stittrich, A.B.,
481 Ament, S.A., Roach, J.C., Brunkow, M.E., et al. (2015). Identification of copy number variants in
482 whole-genome data using Reference Coverage Profiles. Front. Genet. *6*.

483 Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R.K., Lisch, D., Meyers, B.C., Shiu, S.-H., and Jiang, N.
484 (2009). The Functional Role of Pack-MULEs in Rice Inferred from Purifying Selection and Expression
485 Profile. Plant Cell *21*, 25–38.

486 Havecker, E.R., Gao, X., and Voytas, D.F. (2004). The diversity of LTR retrotransposons. Genome Biol.
487 *5*, 225.

488 Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., and Wessler, S.R. (2004). Pack-MULE transposable elements
489 mediate gene evolution in plants. Nature *431*, 569–573.

490 Jiang, N., Ferguson, A.A., Slotkin, R.K., and Lisch, D. (2011). Pack-Mutator–like transposable elements
491 (Pack-MULEs) induce directional modification of genes through biased insertion and DNA
492 acquisition. Proc. Natl. Acad. Sci. *108*, 1537–1542.

493 Juretic, N., Hoen, D.R., Huynh, M.L., Harrison, P.M., and Bureau, T.E. (2005). The evolutionary fate of
494 MULE-mediated duplications of host gene fragments in rice. Genome Res. *15*, 1292–1297.

495 Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C.,
496 Verbsky, M.L., and Richards, E.J. (2003). Arabidopsis MET1 Cytosine Methyltransferase Mutants.
497 Genetics *163*, 1109–1122.

498 Kawasaki, S., and Nitasaka, E. (2004). Characterization of Tpn1 Family in the Japanese Morning Glory:
499 En/Spm-related Transposable Elements Capturing Host Genes. Plant Cell Physiol. *45*, 933–944.

500 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*,
501 357–359.

502 Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation
503 patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

504 Lisch, D. (2013). How important are transposons for plant evolution? Nat. Rev. Genet. *14*, 49–61.

505 Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R.
506 (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell *133*,
507 523–536.

508 Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel,
509 D., Paszkowski, J., and Mathieu, O. (2009). Selective epigenetic control of retrotransposition in
510 Arabidopsis. Nature *461*, 427–430.

511 Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001).
512 Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature
513 *411*, 212–214.

514 Nakagome, M., Solovieva, E., Takahashi, A., Yasue, H., Hirochika, H., and Miyao, A. (2014).
515 Transposon Insertion Finder (TIF): a novel program for detection of de novo transpositions of
516 transposable elements. BMC Bioinformatics *15*, 71.

517 Ralston, E., English, J., and Dooner, H.K. (1989). Chromosome-breaking structure in maize involving a
518 fractured Ac element. Proc. Natl. Acad. Sci. *86*, 9451–9455.

519 Reinders, J., Wulff, B.B.H., Mirouze, M., Marí-Ordóñez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler,
520 G., and Paszkowski, J. (2009). Compromised stability of DNA methylation and transposon
521 immobilization in mosaic Arabidopsis epigenomes. Genes Dev. *23*, 939–950.

522 Saze, H., Scheid, O.M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for
523 epigenetic inheritance during plant gametogenesis. Nat. Genet. *34*, 65–69.

524 Skipper, K.A., Andersen, P.R., Sharma, N., and Mikkelsen, J.G. (2013). DNA transposon-based gene
525 vehicles - scenes from an evolutionary drive. J. Biomed. Sci. *20*, 92.

526 Talbert, L.E., and Chandler, V.L. (1988). Characterization of a highly conserved sequence related to
527 mutator transposable elements in maize. Mol. Biol. Evol. *5*, 519–529.

528 Weil, C.F. (2009). Too many ends: aberrant transposition. Genes Dev. *23*, 1032–1036.

529 Xu, M., Brar, H.K., Grosic, S., Palmer, R.G., and Bhattacharyya, M.K. (2010). Excision of an Active
530 CACTA-Like Transposable Element From DFR2 Causes Variegated Flowers in Soybean [Glycine max
531 (L.) Merr.]. Genetics *184*, 53–63.

532 Yamashita, S., Takano-Shimizu, T., Kitamura, K., Mikami, T., and Kishima, Y. (1999). Resistance to Gap
533 Repair of the Transposon Tam3 in Antirrhinum majus: A Role of the End Regions. Genetics *153*,
534 1899–1908.

535 Yoon, S., Xuan, Z., Makarov, V., Ye, K., and Sebat, J. (2009). Sensitive and accurate detection of copy
536 number variants using read depth of coverage. Genome Res. *19*, 1586–1592.

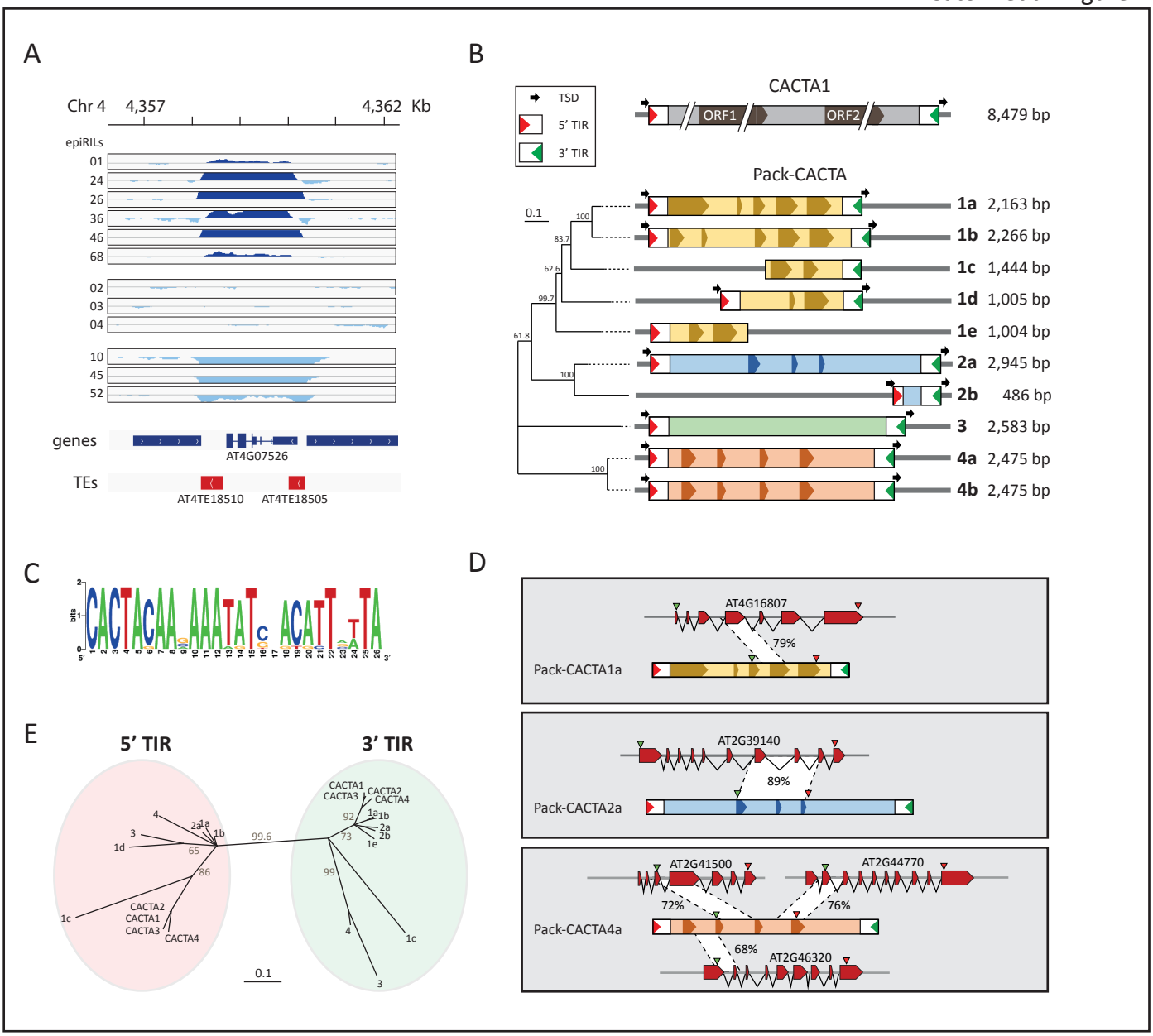
537 Yu, Z., Wright, S.I., and Bureau, T.E. (2000). Mutator-like Elements in Arabidopsis thaliana: Structure,
538 Diversity and Evolution. Genetics *156*, 2019–2031.

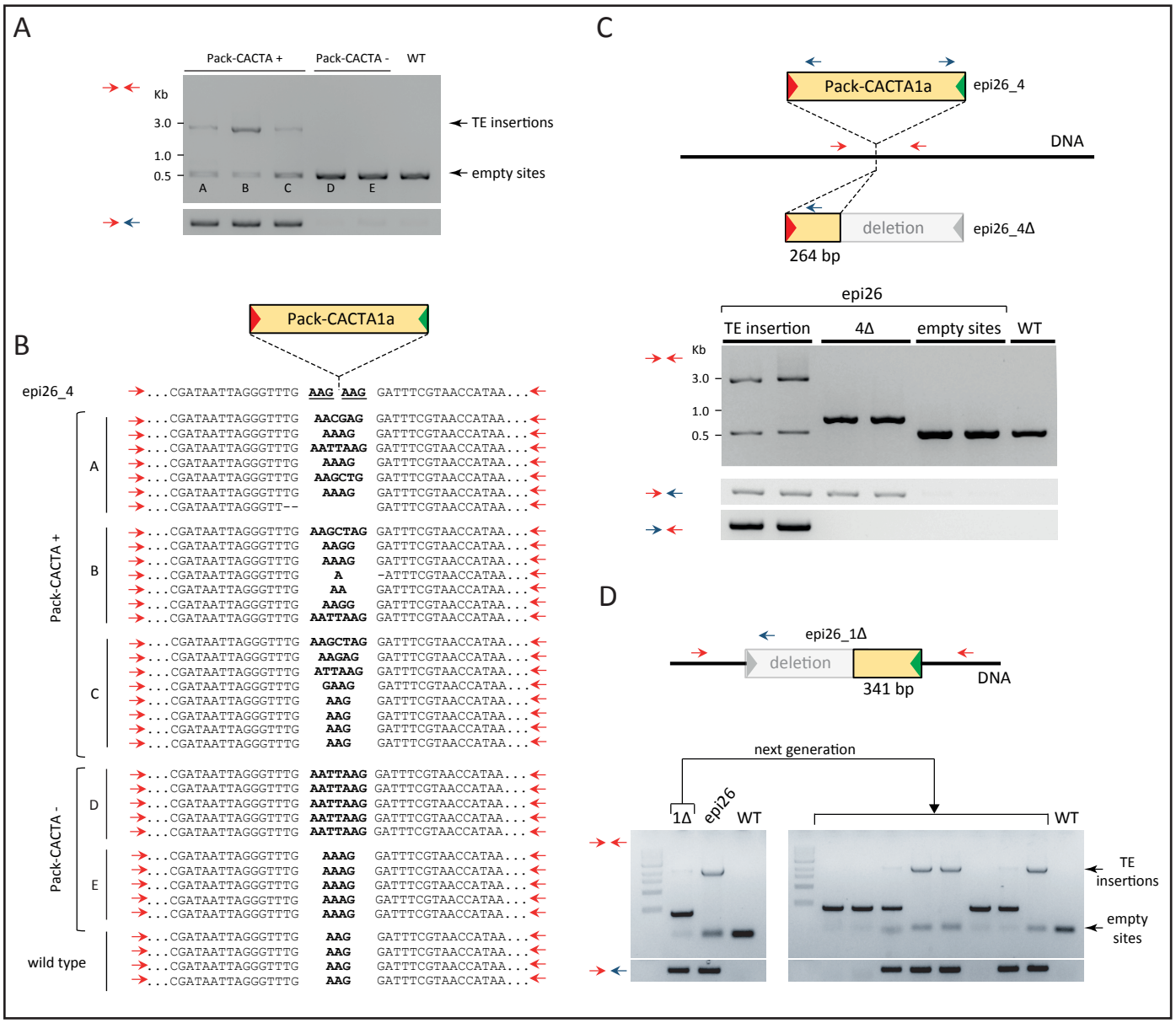
539 Zabala, G., and Vodkin, L.O. (2005). The *wp* Mutation of *Glycine max* Carries a Gene-Fragment-Rich
540 Transposon of the CACTA Superfamily. Plant Cell Online *17*, 2619–2632.

541 Zhang, P., and Spradling, A.C. (1993). Efficient and Dispersed Local P Element Transposition from
542 Drosophila Females. Genetics *133*, 361–373.

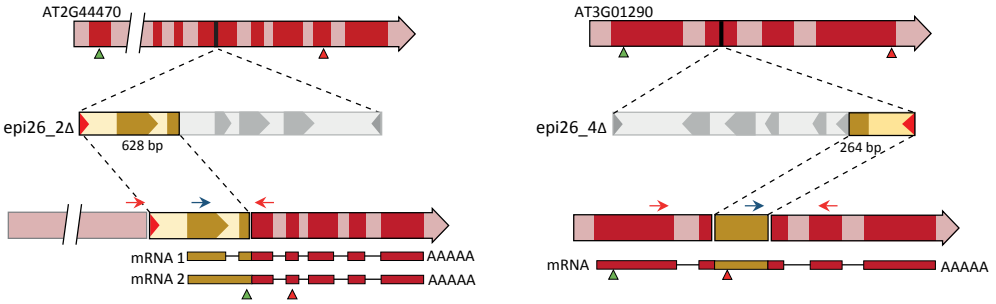
543 Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of
544 Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and
545 transcription. Nat. Genet. *39*, 61–69.

546

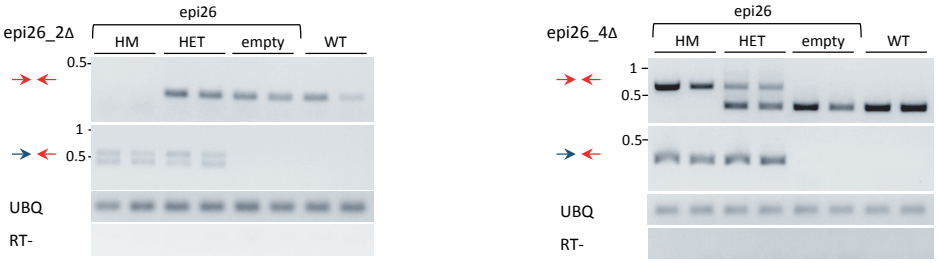

A
Chr 4
4,357
4,362
Kb
epiRILs
01
24
26
36
46
68
02
03
04
10
45
52
genes
AT4G07526
TEs
AT4TE18510
AT4TE18505
B
TSD
5' TIR
3' TIR
CACTA1
ORF1
ORF2
8,479 bp
Pack-CACTA
0.1
1a 2,163 bp
1b 2,266 bp
1c 1,444 bp
1d 1,005 bp
1e 1,004 bp
2a 2,945 bp
2b 486 bp
3 2,583 bp
4a 2,475 bp
4b 2,475 bp
C
bits
D
AT4G16807
79%
Pack-CACTA1a
AT2G39140
89%
Pack-CACTA2a
AT2G41500
AT2G44770
72%
76%
Pack-CACTA4a
68%
AT2G46320
E
5' TIR
3' TIR
99.6
65
86
0.1

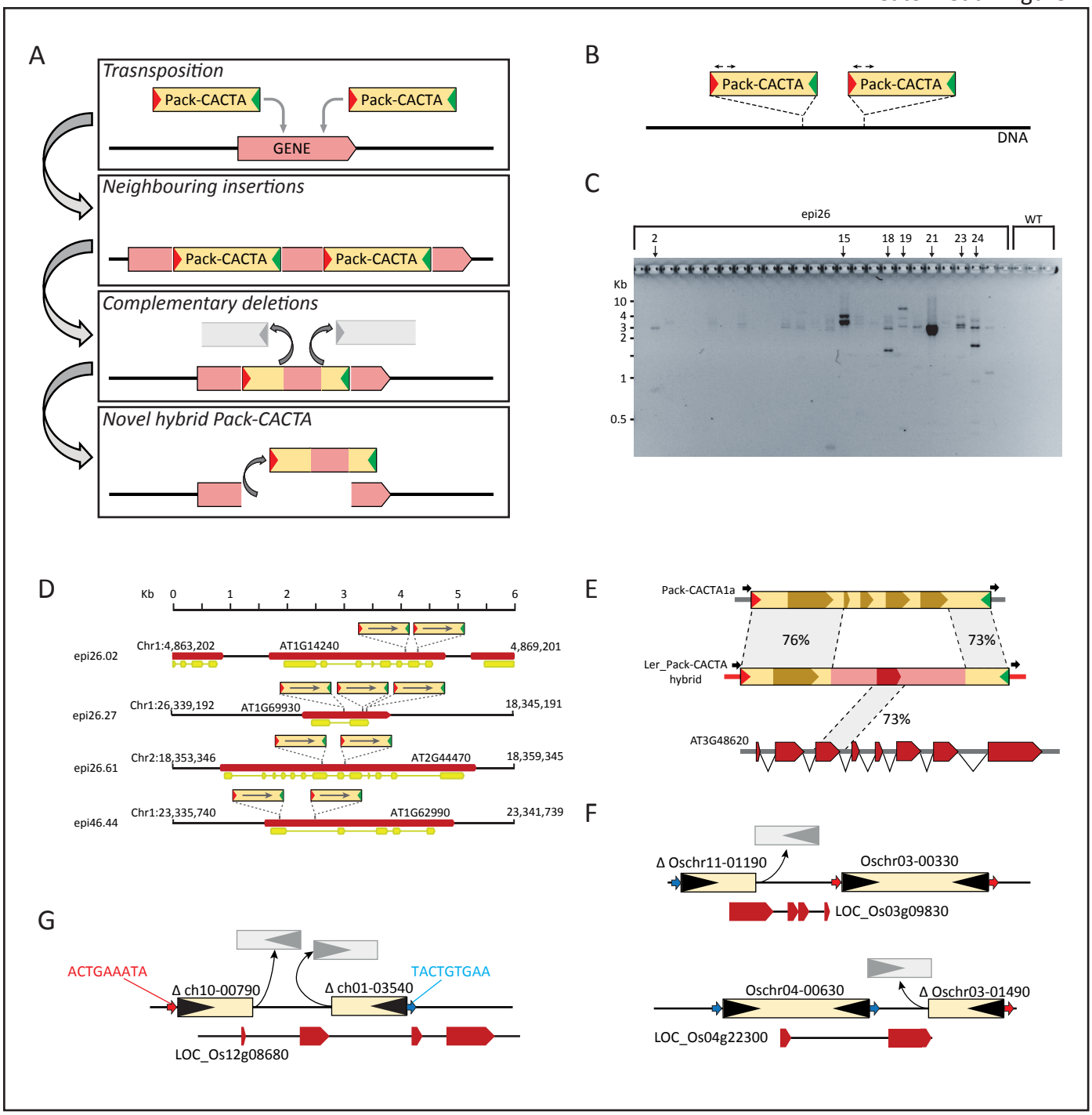

A
Pack-CACTA +
Pack-CACTA -
WT
Kb
3.0
1.0
0.5
A
B
C
D
E
TE insertions
empty sites
B
Pack-CACTA1a
epi26_4
Pack-CACTA +
Pack-CACTA -
wild type
C
Pack-CACTA1a
epi26_4
DNA
deletion
epi26_4Δ
264 bp
epi26
TE insertion
4Δ
empty sites
WT
D
epi26_1Δ
deletion
341 bp
DNA
next generation
1Δ
epi26
WT
WT
TE insertions
empty sites

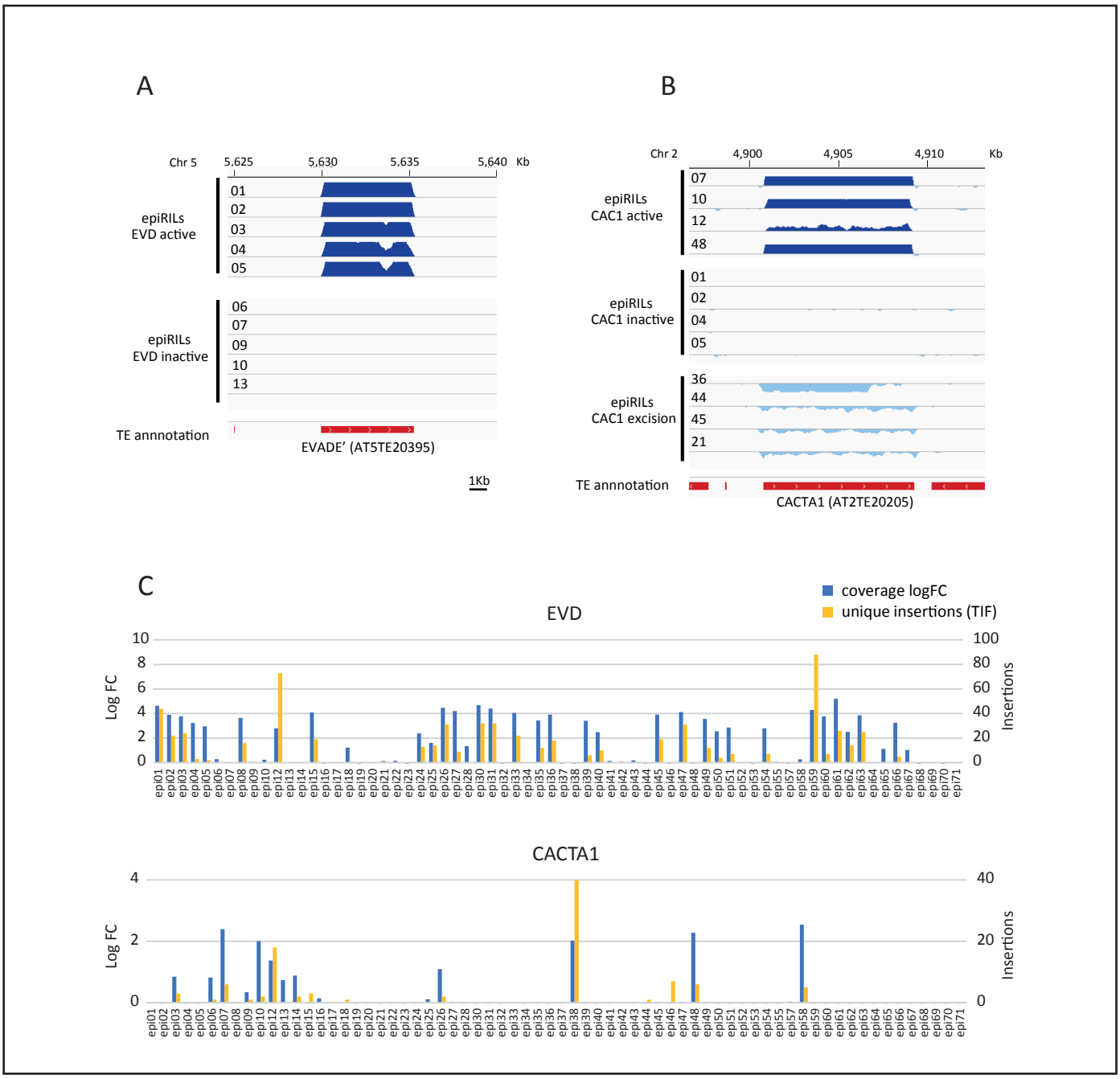
A

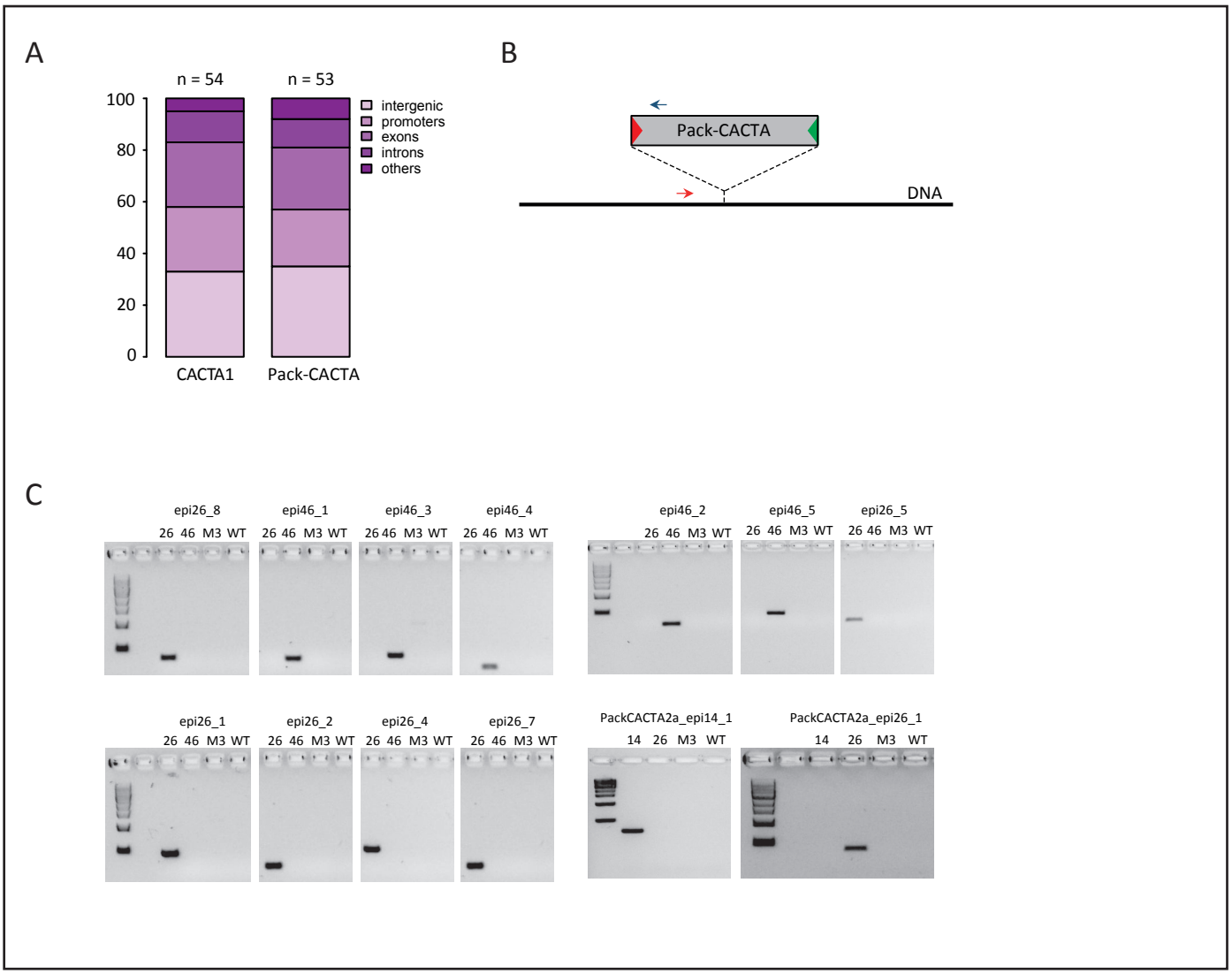

AT2G44470
epi26_2Δ
628 bp
mRNA 1
mRNA 2
AAAAA
AAAAA
AT3G01290
epi26_4Δ
264 bp
mRNA
AAAAA

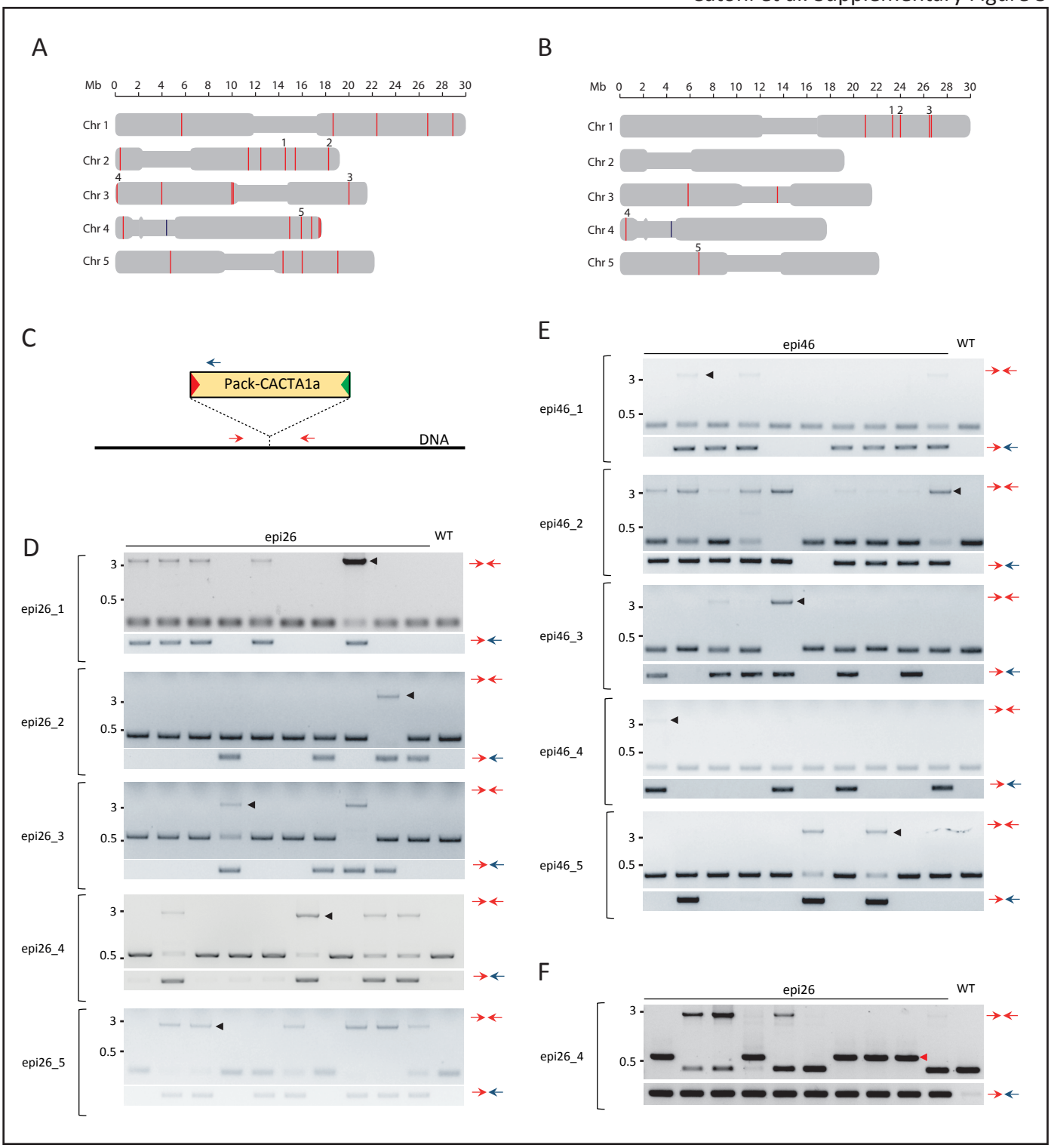
B

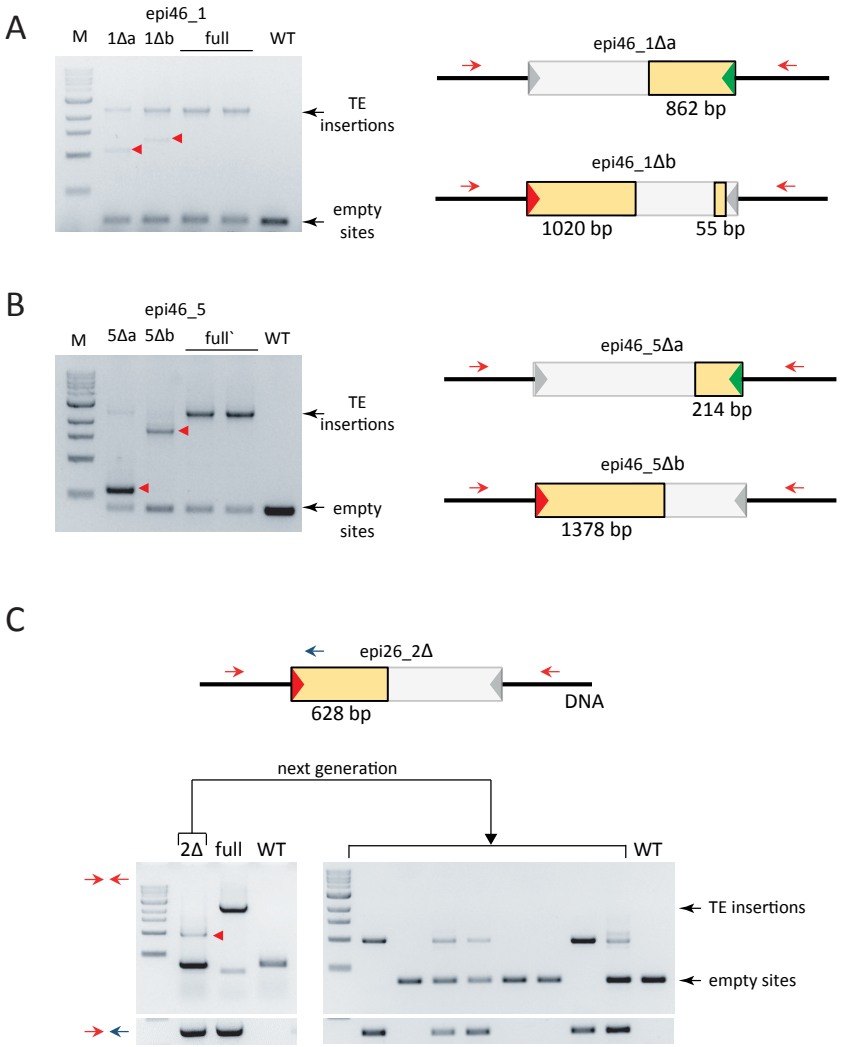

epi26
epi26_2Δ
HM
HET
empty
WT
0.5
1
0.5
UBQ
RT-
epi26
epi26_4Δ
HM
HET
empty
WT
1
0.5
0.5
UBQ
RT-


A
Trasnsposition
Pack-CACTA
Pack-CACTA
GENE
Neighbouring insertions
Pack-CACTA
Pack-CACTA
Complementary deletions
Novel hybrid Pack-CACTA
B
Pack-CACTA
Pack-CACTA
DNA
C
epi26
WT
2
15
18 19
21
23 24
Kb
10
4
3
2
1
0.5
D
Kb 0 1 2 3 4 5 6
epi26.02
Chr1:4,863,202
AT1G14240
4,869,201
epi26.27
Chr1:26,339,192
AT1G69930
18,345,191
epi26.61
Chr2:18,353,346
AT2G44470
18,359,345
epi46.44
Chr1:23,335,740
AT1G62990
23,341,739
E
Pack-CACTA1a
76%
73%
Ler_Pack-CACTA hybrid
73%
AT3G48620
F
Δ Oschr11-01190
Oschr03-00330
LOC_Os03g09830
Oschr04-00630
Δ Oschr03-01490
LOC_Os04g22300
G
ACTGAAATA
Δ ch10-00790
Δ ch01-03540
TACTGTGAA
LOC_Os12g08680

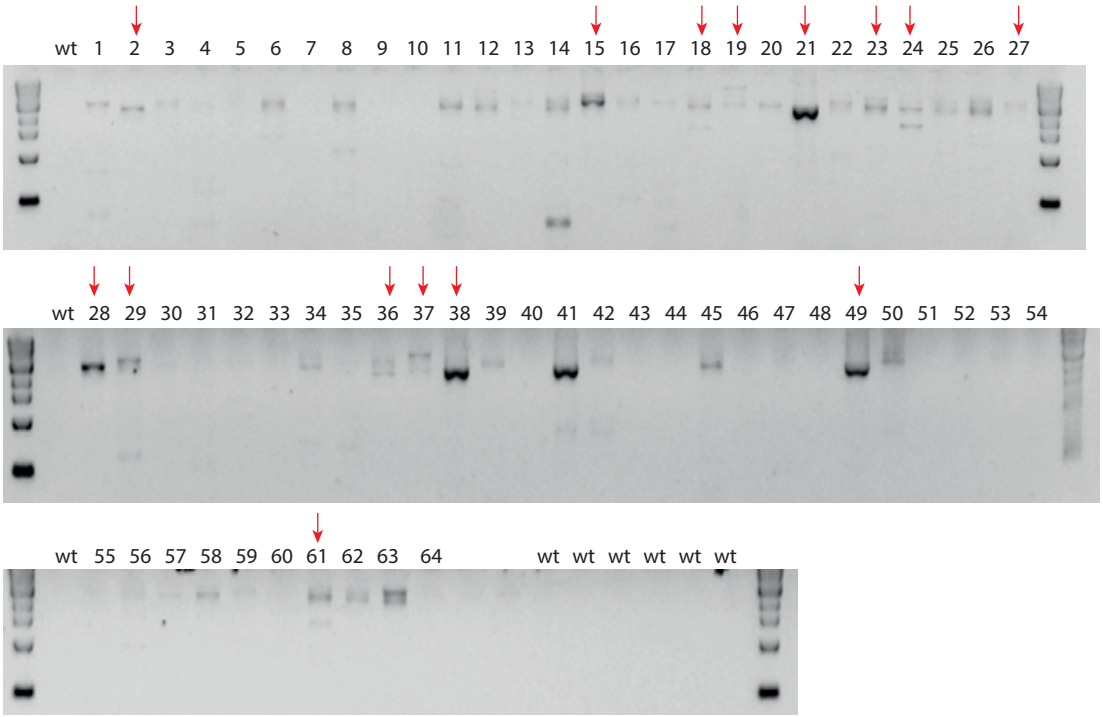

A
Chr 5
5,625
5,630
5,635
5,640
Kb
epiRILs
EVD active
01
02
03
04
05
epiRILs
EVD inactive
06
07
09
10
13
TE annnotation
EVADE' (AT5TE20395)
1Kb
B
Chr 2
4,900
4,905
4,910
Kb
epiRILs
CAC1 active
07
10
12
48
epiRILs
CAC1 inactive
01
02
04
05
epiRILs
CAC1 excision
36
44
45
21
TE annnotation
CACTA1 (AT2TE20205)
C
EVD
coverage logFC
unique insertions (TIF)
Log FC
Insertions
CACTA1
Log FC
Insertions


A
n = 54
n = 53
100
80
60
40
20
0
intergenic
promoters
exons
introns
others
CACTA1
Pack-CACTA
B
Pack-CACTA
DNA
C
epi26_8
26 46 M3 WT
epi46_1
26 46 M3 WT
epi46_3
26 46 M3 WT
epi46_4
26 46 M3 WT
epi46_2
26 46 M3 WT
epi46_5
26 46 M3 WT
epi26_5
26 46 M3 WT
epi26_1
26 46 M3 WT
epi26_2
26 46 M3 WT
epi26_4
26 46 M3 WT
epi26_7
26 46 M3 WT
PackCACTA2a_epi14_1
14 26 M3 WT
PackCACTA2a_epi26_1
14 26 M3 WT

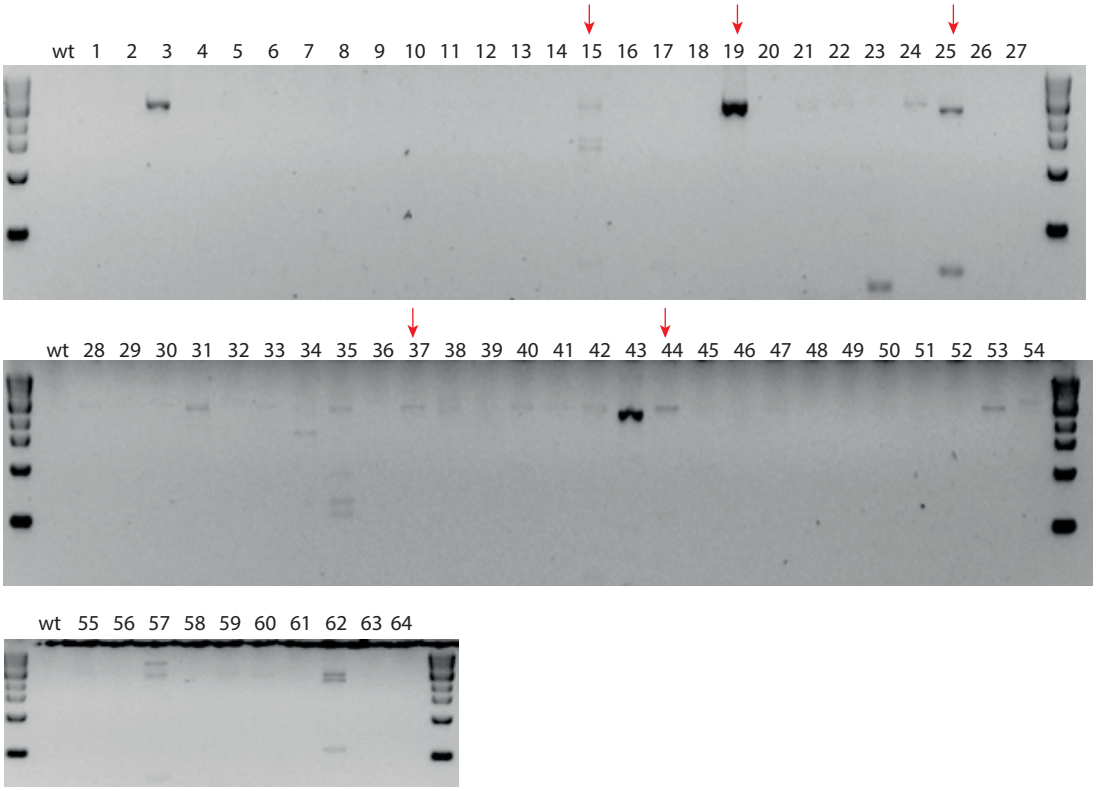

A
Mb 0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
B
C
Pack-CACTA1a
DNA
D
epi26
WT
epi26_1
epi26_2
epi26_3
epi26_4
epi26_5
3
0.5
E
epi46
epi46_1
epi46_2
epi46_3
epi46_4
epi46_5
F

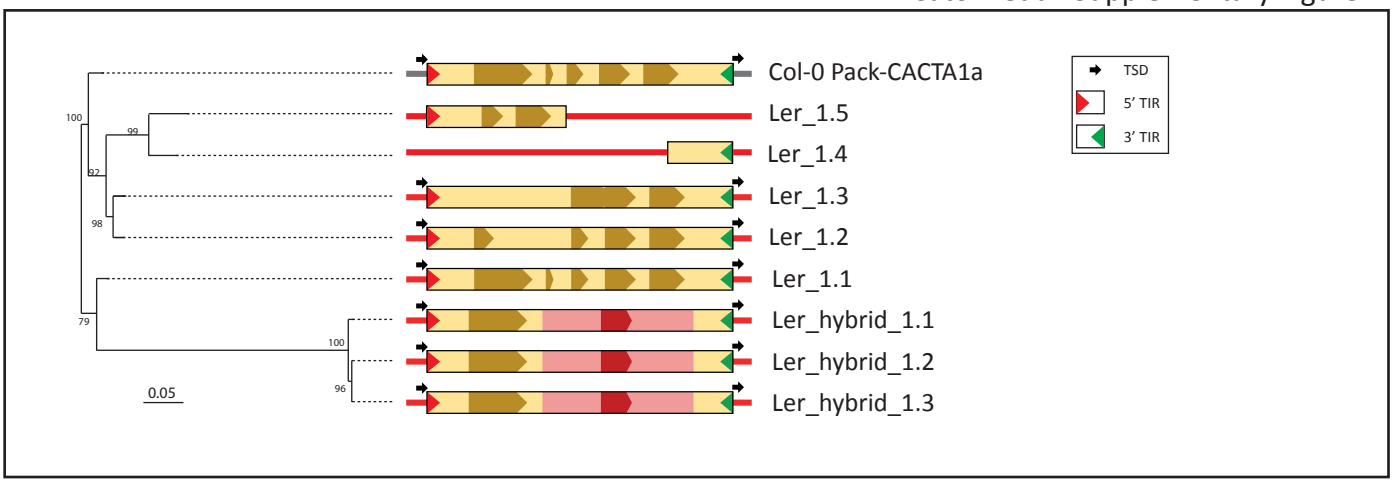

A
epi46_1
M
1Δa
1Δb
full
WT
TE insertions
empty sites
epi46_1Δa
862 bp
epi46_1Δb
1020 bp
55 bp
B
epi46_5
M
5Δa
5Δb
full`
WT
TE insertions
empty sites
epi46_5Δa
214 bp
epi46_5Δb
1378 bp
C
epi26_2Δ
628 bp
DNA
next generation
2Δ
full
WT
WT
TE insertions
empty sites

A


wt 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
wt 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
wt 55 56 57 58 59 60 61 62 63 64 wt wt wt wt wt wt

B


wt 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
wt 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
wt 55 56 57 58 59 60 61 62 63 64

Plant	locus	coordinates	orientation	location
epi26.02a	150 bp	Chr1:26795735-26795884	F	AT1G71030
epi26.02b	209 bp	Chr1:4867289-4867497	F	AT1G14240
epi26.15a	592 bp	Chr1:7331830-7332421	R	AT1G20990
epi26.15b	1041 bp	Chr1:3907652-3908629	F	AT1G11610
epi26.19	> 1618 bp	Chr1:1848083-1849700...	R	AT1G06090
epi26.21	255 bp	Chr3:19478022-19478276	R	AT3G52527
epi26.23	812 bp	Chr1:7332419-7333230	R	intergenic
epi26.24	Δ 1835 bp, 1323 bp	Chr1:10576415-10577737	F	intergenic
epi26.27a	53 bp	Chr1:26337760-26337812	R	AT1G69930
epi26.27b	342 bp	Chr1:26337810-26338152	R	AT1G69930
epi26.27c	38 bp	Chr3:4979127-4979164	F	AT3G14820
epi26.27d	231 bp	Chr5:23417812-23418042	F	intergenic
epi26.28	508 bp	Chr4:8280700-8281207	F	AT4G14370
epi26.29	> 947 bp	Chr2:...7892207-7893153	F	AT2G18150
epi26.36	517 bp	Chr2:17585967-17586483	R	intergenic
epi26.37	> 988 bp	Chr3:20273633-20274620...	R	AT3G54770
epi26.38a	166 bp	Chr5:6584099-6584264	F	intergenic
epi26.38b	768 bp	Chr2:10944660-10945427	F	intergenic
epi26.49	Δ 147 bp, 531 bp	Chr2:9234975-9235505	F	intergenic
epi26.61a	360 bp	Chr3:1145525-1145884	R	AT3G04330
epi26.61b	408 bp	Chr2:18355972-18356379	F	AT2G44470
epi46.15	434 bp	Chr3:2131389-2131822	R	intergenic
epi46.19	537 bp	Chr5:6566750-6567286	R	AT5G19470
epi46.25	184 bp	Chr5:351068-351251	F	AT5G01900
epi46.37	435 bp	Chr1:17660330-17660765	R	AT1G47915
epi46.44a	638 bp	Chr1:2333762-23338258	F	AT1G62990
epi46.44b	Δ 1747 bp, 2137 bp	Chr1:21016481-21018611	F	AT1G56150
epi26.18	1182 bp	Chr5:4025909-4027090		AT5G12420

Catoni et al. Supplementary Figure 7