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

Marco Catoni*, Thomas Jonesman, Elisa Cerruti and Jerzy Paszkowski*

Affiliations

The Sainsbury Laboratory, University of Cambridge, Cambridge, 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.

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).

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

27 most active, with an average 11.8-fold coverage increase (Supplementary Table 2). Surprisingly, the
28 second most active element, with a 6.2-fold increase, had an unusually complex structure of two
29 short terminal DNA stretches of 451 bp (AT4TE18505) and 337 bp (AT4TE18510) (annotated as a
30 member of the ATENSPM3 family of DNA transposons) separated by a sequence annotated as gene
31 AT4G07526 of unknown function (Figure 1A). The putative product of this gene showed no similarity
32 to any known transposase and the overall transposon structure, residing on chromosome 4,
33 resembled a non-autonomous Pack-TYPE element.

34 AT4TE18505 and AT4TE18510 contain TIRs beginning with 8 base pairs (CACTACAA) that are also a
35 feature of the CACTA1 autonomous transposon (Miura et al., 2001), which is active in the epiRILs
36 (Reinders et al., 2009) (Supplementary Table 2). A blast search with 150 bp of both terminal
37 sequences identified nine additional elements in the Col-0 genome with similar structures and
38 CACTA1-like termini (Figure 1B). These included conserved terminal sequences of 26 bp (Figure 1C)
39 and a 3-bp target site duplication (TSD) (Supplementary Table 3). Their 150-bp terminal sequence
40 identities ranged from 62 to 95.3% (Supplementary Table 4). Five of the elements shared related
41 DNA sequences located between the CACTA1-like termini, while the remaining five elements carried
42 sequences of different chromosomal origin (Figure 1B and D). Interestingly, not all of the captured
43 sequences displayed similarities to Arabidopsis genes (Figure 1D) and DNA sequence databases did
44 not reveal their origins. The TIR sub-terminal sequences of CACTA family transposons differ slightly
45 between the 5' and the 3' ends (Frey et al., 1990). These differences were also present in the
46 possibly active Pack-TYPE elements (Figure 1E) we discovered. Therefore, we called this family Pack-
47 CACTA, which in the Arabidopsis Col-0 accession consists of 10 TEs divided up, according to their
48 sequence similarities, into four groups (Figure 1B).

49 We found 50 new insertions of Pack-CACTA1a and 3 new insertions of Pack-CACTA2a in 8 and 3
50 epiRILs, respectively (Supplementary Table 5). In line epi26, Pack-CACTA1a and Pack-CACTA2a
51 appeared mobile (Supplementary Table 5). New transposed copies were found mostly within

52 euchromatic chromosomal regions (65%), resembling the insertion preference of CACTA1
53 (Supplementary Figure 2A). We validated new insertions of Pack-CACTA1a by locus-specific PCR,
54 choosing 11 random sites in epi26 and epi46 and two sites with Pack-CACTA2a insertions in epi14
55 and epi26 (Supplementary Figure 2B, C and Supplementary Table 6). Thus, both Pack-CACTA1a and
56 Pack-CACTA2a were currently transposing and generating TSD of three nucleotides (Supplementary
57 Data 1).

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

68 Unexpectedly, in some plants of epi26 locus 4 (epi26_4) (Supplementary Figure 3F) we found a 264-
69 bp fragment of the CACTA1a 5' sequence flanked by 3 bp of TDS (Figure 2C and Supplementary Data
70 3). Ten epi26 plants with a Pack-CACTA1a insertion at this locus contained either the entire Pack-
71 CACTA1a or its deletion derivative (Supplementary Figure 3F). Therefore, it is most probable that the
72 deletion resulted from aberrant excision of a full copy from this location. Importantly, we did not
73 detect somatic "reversions" to a wild type-like locus in plants homozygous for the deleted version of
74 Pack-CACTA1a. The fragmented element was apparently immobilized at this new location (Figure
75 2C).

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

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

92 The sequences of remaining fragments of the transposon showed that deletions occurred at
93 approximately equal frequencies at the 3' (3 deletions) and 5' terminals (4 deletions) (Figure 2C,D,
94 Supplementary Figure 4, Supplementary Data 5). This observation suggested that incorporation of
95 chromosomal DNA into a novel Pack-CACTA element requires the presence of two neighbouring
96 transposons inserted in the direct orientation. If such a pair of Pack-Type transposons, separated by
97 a short stretch of chromosomal DNA, undergoes complementary deletions of each of the elements,
98 a "hybrid" Pack-CACTA encompassing chromosomal DNA within the respective 5' and 3' terminals of
99 the former transposons would be generated (Figure 4A). To examine whether Pack-CACTA1a
100 integrates frequently at closely linked chromosomal locations, we designed a PCR-based screen to

101 recover chromosomal DNA stretches residing between putative neighbouring new inserts of Pack-
102 CACTA1a (Figure 4B). Obviously, this approach can only detect insertions separated by relatively
103 short stretches of chromosomal DNA. Therefore, it was surprising that numerous amplifications of
104 chromosomal DNA fragments were observed in the 128 DNA samples of epi26 and epi46 used
105 previously for the Pack-CACTA1a deletion screen (Figure 4C and Supplementary Figure 5). The
106 sequences of 28 different PCR products from 20 plants (Supplementary Figure 6 and Supplementary
107 Data 6, 7) all contained fragments of Arabidopsis genomic DNA of various sizes (38 to 2,137 bp)
108 between two Pack-CACTA1a copies. Remarkably, 27 PCR products revealed insertions of Pack-
109 CACTA1a as tandems in direct orientation and only one inversion (Supplementary figure 6). The
110 frequent formation by newly inserted copies of Pack-CACTA1a of direct repeats interspaced by short
111 stretches of chromosomal DNA, combined with the observed formation of terminal deletion
112 derivatives of the transposon, suggest that pairs of Pack-CACTA1a may form new “hybrid” elements.
113 Indeed, we recovered three deletion derivatives of Pack-CACTA1a among the 27 direct transposon
114 repeats (Supplementary Figure 6). Importantly, most neighbouring Pack-CACTA1a insertions were
115 separated by genic sequences (21 out of 28, Supplementary Figure 6); in one case three Pack-
116 CACTA1a insertions were located in one gene (Figure 4D).

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

122 As an alternative, we searched for structural variants of Pack-CACTA1a elements in the newly
123 assembled genome of the Arabidopsis Ler-0 accession and retrieved eight analogous elements
124 (Supplementary Figure 7 and Supplementary Table 7). Of these, the sequences of three full-length
125 elements and two deletion derivatives were closely related to Pack-CACTA1a of Col-0, with sequence

126 identities of 73 to 90% (Supplementary Table 7). The sequences of the remaining three copies of Ler-
127 0 Pack-CACTA elements resembled Pack-CACTA1a only in the first 992 bp and the last 325 bp with
128 76% and 72% identities, respectively. Remarkably, the internal 1,181 bp showed no similarity to the
129 "parental" element but contained a 132-bp stretch with 73% identity to the second exon of gene
130 AT3G48620 encoding a putative protein of the outer membrane family (OMP85) (Figure 4E). This
131 novel structure of a Ler-0 relative of Pack-CACTA1a suggests that a Pack-CACTA transposon in this
132 Arabidopsis accession could capture and move chromosomal DNA by a mechanism compatible with
133 that we recorded in real time for Pack-CACTA1a transposition in Col-0.

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

151 A massive contribution of Pack-TYPE transposon activity to the evolution of plant genes and
152 genomes was deduced from analyses of rice, Brassica and soybean (Alix et al., 2008; Hanada et al.,
153 2009; Jiang et al., 2004; Jiang et al., 2011; Zabala and Vodkin, 2005). It was postulated that repair of
154 nicks and gaps created at the transposon insertions sites or arising due to the element structures
155 might be responsible for the acquisition of chromosomal DNA (Bennetzen and Springer, 1994; Engels
156 et al., 1990; Yamashita et al., 1999). Such DNA-repair mechanisms would operate prior transposon
157 relocation and the newly generated elements would than move to different chromosomal locations
158 only after incorporation of chromosomal DNA. Here, when studying transposition properties of Pack-
159 CACTA1a elements *in vivo*, we found that acquisition of chromosomal DNA seems to be tightly linked
160 to the transposition. Pack-CACTA1a transposition has a marked tendency to produce closely spaced
161 repeats in direct orientation. Subsequently, complementary aberrant excisions of the elements
162 within these repeats, leaving behind one of the two TIRs, generate a novel "hybrid" Pack-CACTA that
163 incorporates the chromosomal DNA that separated the two neighbouring insertions (Figure 4A).

164 This quite simple and efficient mechanism of chromosomal DNA acquisition relies on the
165 combination of known characteristics of particular DNA transposons. For example, maize Ac/Ds, P-
166 element of *Drosophila,* Tc1 transposon in *Caenorhabditis* and *Sleeping Beauty* in mouse tend to
167 transpose to nearby locations, termed "local hopping" (Bancroft and Dean, 1993; Carlson et al.,
168 2003; Fischer et al., 2003; Skipper et al., 2013; Zhang and Spradling, 1993). However, the orientation
169 of elements in such integration arrays are random and not preferentially direct, as observed for
170 Pack-CACTA1a. Incomplete or aberrant excision is known for CACTA-like transposable elements in
171 soybean (Xu et al., 2010), while maize chromosomal aberrations, like breakages and fusions, were
172 previously associated with the transposition of TIRs belonging to two different neighbouring Ac/Ds
173 elements (Ralston et al., 1989). It is interesting that Pack-CACTA1a combines both properties, i.e.
174 efficient formation of closely linked tandem insertions in direct orientation and frequent aberrant
175 excisions that lead to complementary deletion of one of the two TIRs of each pair. The remaining

176 terminals of the neighbouring transposons encompass a new, single element incorporating the
177 intervening chromosomal DNA.

178 The incorporation of gene fragments and their duplication by transposition to new genomic
179 positions has been suggested for certain TE families, including MULEs, CACTAs and HELITRONs (Lisch,
180 2013). This “transduplication” process seems to be very frequent in rice and can be clearly assigned
181 to the historical activity of Pack-MULEs (Juretic et al., 2005). Importantly, transduplicated genes
182 seem to be under purifying selection (Hanada et al., 2009) or serve regulatory functions (Juretic et
183 al., 2005). This implies that Pack-TYPE transposon mobilization may have a direct and profound
184 influence on gene evolution, generating new gene functions or regulatory activities by shuffling parts
185 of various coding regions across the genome. Taking into account the high rates of nearby
186 directional insertion and aberrant excision observed here for the Pack-CACTA1a element, the “TE-
187 assisted” origin of genes may be strongly underestimated, especially in plants with genomes larger
188 than Arabidopsis.

189 # Methods

190 **Plant material**

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

195 **DNA extraction and library preparation**

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

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

207 The DNA libraries were sequenced with 2 x 76-bp paired-end reads with a minimum of 10x coverage
208 (24x on average) on an Illumina NextSeq 500 using the High-Output Flow Cell configuration. The raw
209 reads were trimmed using Trimmomatic to remove adapter sequences. Reads with an averaged
210 value of at least 15 in a 4-nt window were trimmed from both ends. After trimming, reads pairs with
211 at least one mate shorter that 36 bp were discarded. The remaining sequences (on average 95% of
212 raw reads) were aligned with bowtie2 (Langmead and Salzberg, 2012) with the –no mixed and –non

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

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

251 **RT PCR**

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

258 **BLAST search**

259 BLAST analyses were carried out using the NCBI web interface
260 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or the locally installed BLAST+ v2.2.25 (Camacho et al.,
261 2009), depending on the analysis. When the local version was used, blastn was run with the options
262 –evalue 1e-5 and -max_target_seqs 100 where not specified otherwise.

263 **Pack-CACTA analysis**

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

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

275 **Rice Pack-MULE analysis**

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

285 **Accession numbers**

286 Sequencing data have been deposited in Gene Expression Omnibus under the accession number
287 (sequencing data are under submission process).

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.

Acknowledgements

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

295 # Figure legends

296

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

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

303 B) Structure of Pack-CACTA family. Four groups of elements are marked in different colours with
304 exons represented by a darker hue. The phylogenetic tree was obtained using alignment of full-
305 length sequences. Numbers at each node indicate bootstrap support values of 1000 replications. The
306 5' terminal inverted repeat (5' TIR) is marked as a red triangle and 3' TIR as a green triangle. Target
307 site duplications of 3 bp (TSD) are marked as black arrows.

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

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

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

315

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

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

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

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

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

332

333

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

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

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

344

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

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

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

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

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

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

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

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

367

368 ## Supplementary Figure legends

369 **Supplementary Figure 1**

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

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

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

382

383 **Supplementary Figure 2**

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

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

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

390

391 **Supplementary Figure 3**

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

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

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

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

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

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

404 **Supplementary Figure 4**

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

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

412 **Supplementary Figure 5**

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

418 **Supplementary Figure 6**

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

425 **Supplementary Figure 7**

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

430 ## References

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

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

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

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

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

442 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L.
443 (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

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

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

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

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

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

456 Fischer, S.E.J., Wienholds, E., and Plasterk, R.H.A. (2003). Continuous exchange of sequence
457 information between dispersed Tc1 transposons in the Caenorhabditis elegans genome. Genetics
458 *164*, 127–134.

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

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

464 Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R.K., Lisch, D., Meyers, B.C., Shiu, S.-H., and Jiang, N.
465 (2009). The Functional Role of Pack-MULEs in Rice Inferred from Purifying Selection and Expression
466 Profile. Plant Cell *21*, 25–38.

467 Havecker, E.R., Gao, X., and Voytas, D.F. (2004). The diversity of LTR retrotransposons. Genome Biol.
468 *5*, 225.

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

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

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

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

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

481 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*,
482 357–359.

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

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

486 Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R.
487 (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell *133*,
488 523–536.

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

492 Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001).
493 Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature
494 *411*, 212–214.

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

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

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

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

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

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

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

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

513 Yamashita, S., Takano-Shimizu, T., Kitamura, K., Mikami, T., and Kishima, Y. (1999). Resistance to Gap
514 Repair of the Transposon Tam3 in Antirrhinum majus: A Role of the End Regions. Genetics *153*,
515 1899–1908.

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

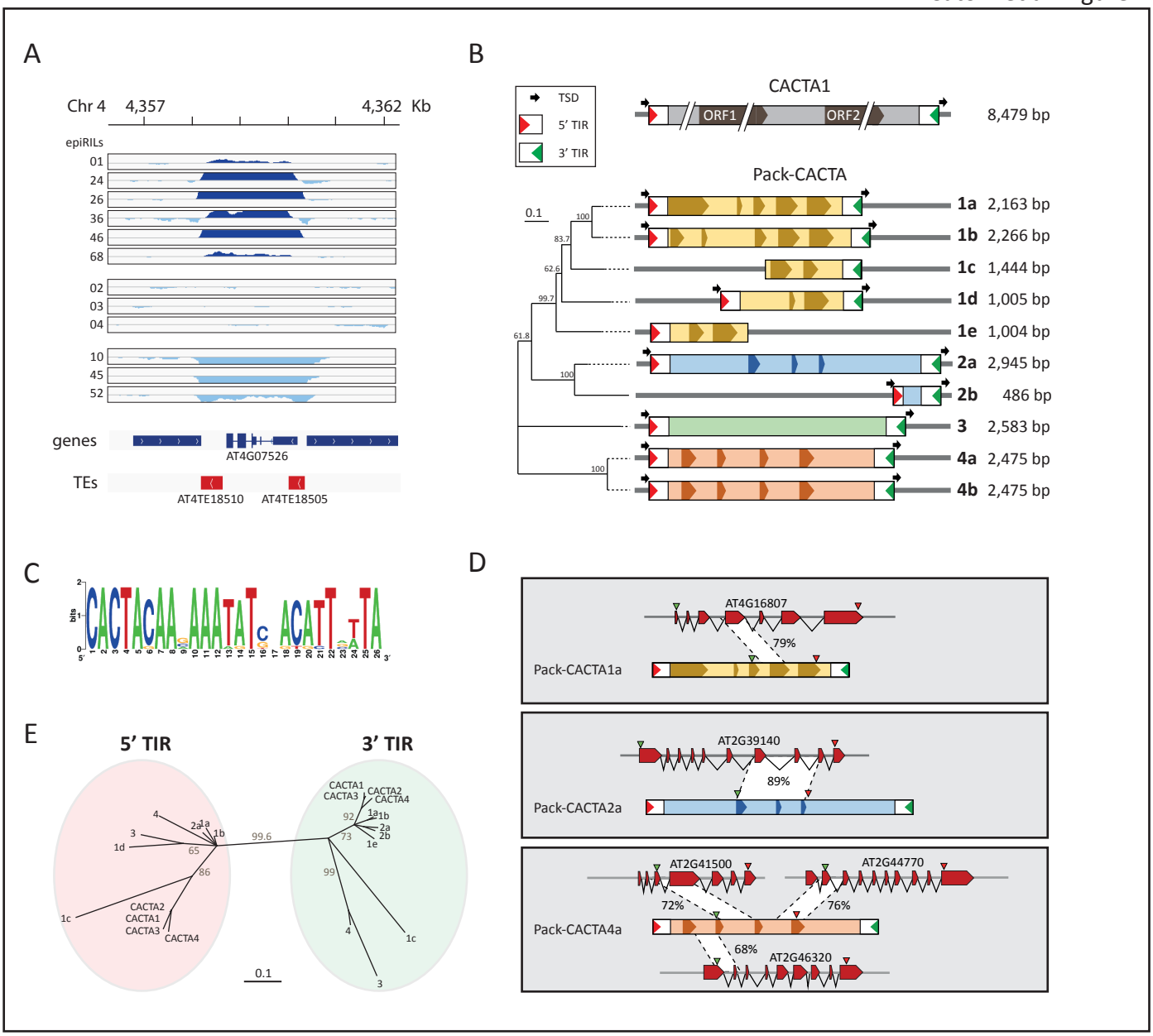
518 Yu, Z., Wright, S.I., and Bureau, T.E. (2000). Mutator-like Elements in Arabidopsis thaliana: Structure,
519 Diversity and Evolution. Genetics *156*, 2019–2031.

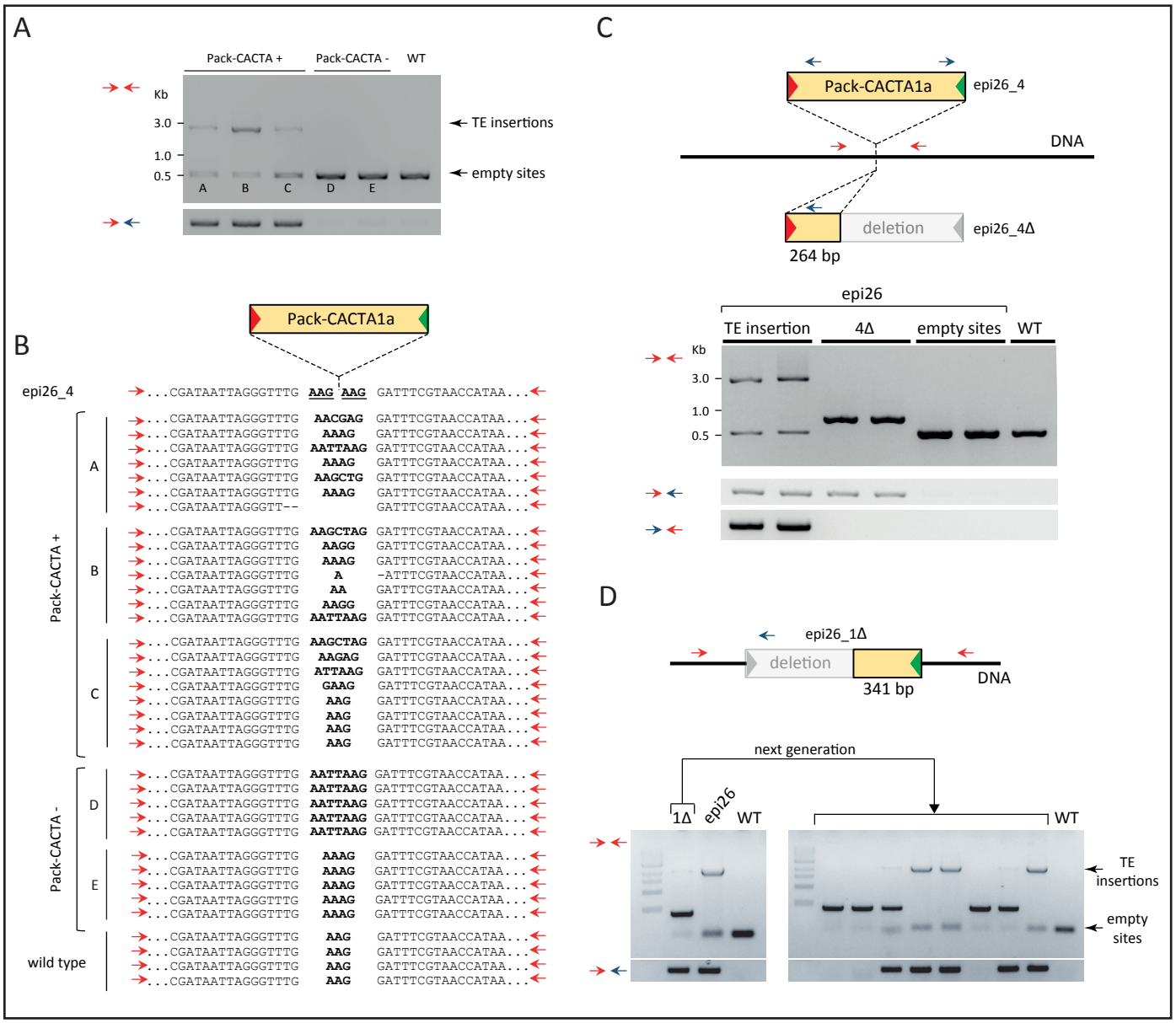
520 Zabala, G., and Vodkin, L.O. (2005). The *wp* Mutation of *Glycine max* Carries a Gene-Fragment-Rich
521 Transposon of the CACTA Superfamily. Plant Cell Online *17*, 2619–2632.

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

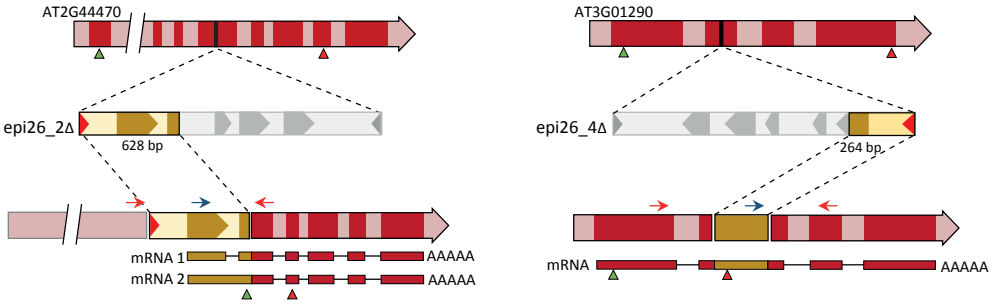
524 Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of
525 Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and
526 transcription. Nat. Genet. *39*, 61–69.

527

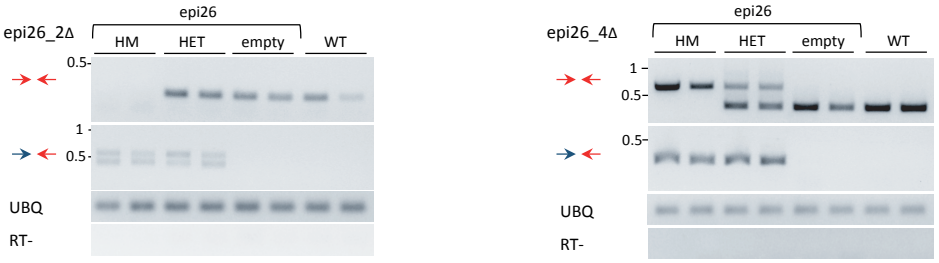

A
Chr 4
4,357
4,362
Kb
epiRILs
01
24
26
36
46
68
02
03
04
10
45
52
genes
AT4G07526
TEs
AT4TE18510
AT4TE18505
B
TSD
5' TIR
3' TIR
CACTA1
ORF1
ORF2
8,479 bp
Pack-CACTA
0.1
1a 2,163 bp
1b 2,266 bp
1c 1,444 bp
1d 1,005 bp
1e 1,004 bp
2a 2,945 bp
2b 486 bp
3 2,583 bp
4a 2,475 bp
4b 2,475 bp
C
bits
D
AT4G16807
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

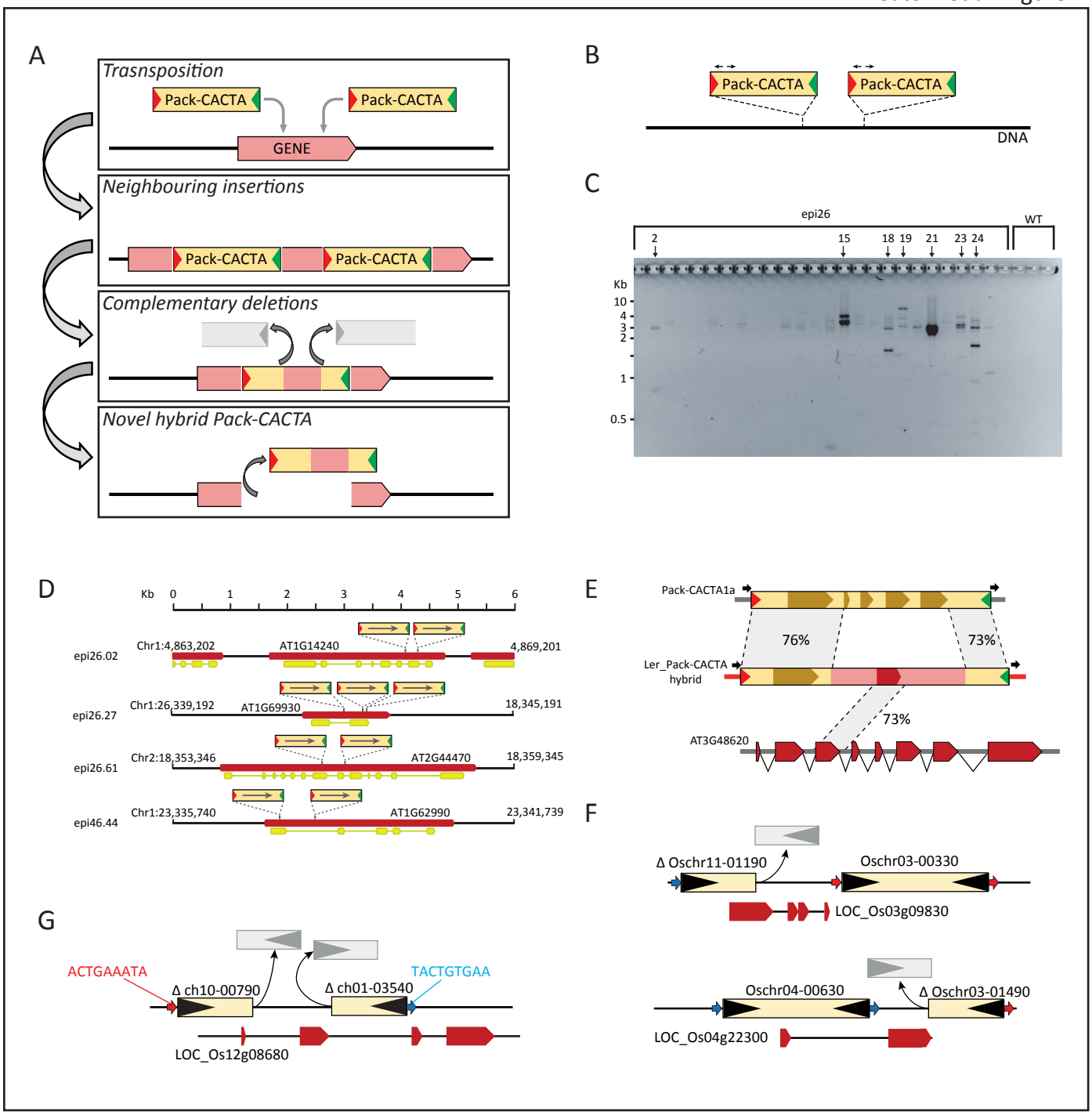

A
Pack-CACTA +
Pack-CACTA -
WT
Kb
3.0
1.0
0.5
A
B
C
D
E
TE insertions
empty sites
B
Pack-CACTA1a
epi26_4
Pack-CACTA +
Pack-CACTA -
wild type
C
Pack-CACTA1a
epi26_4
DNA
deletion
epi26_4Δ
264 bp
epi26
TE insertion
4Δ
empty sites
WT
D
epi26_1Δ
deletion
341 bp
DNA
next generation
1Δ
epi26
WT
WT
TE insertions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

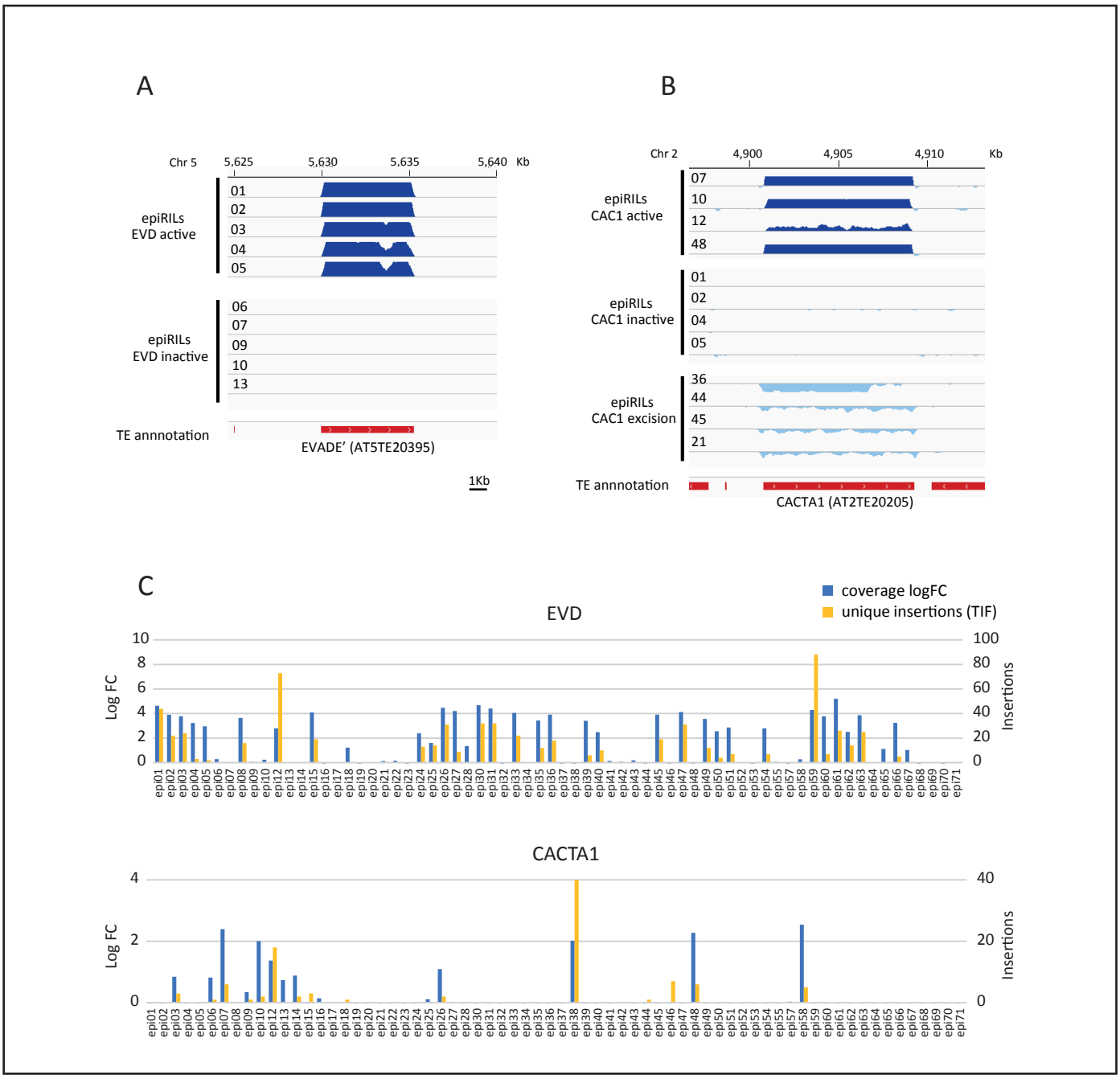
A

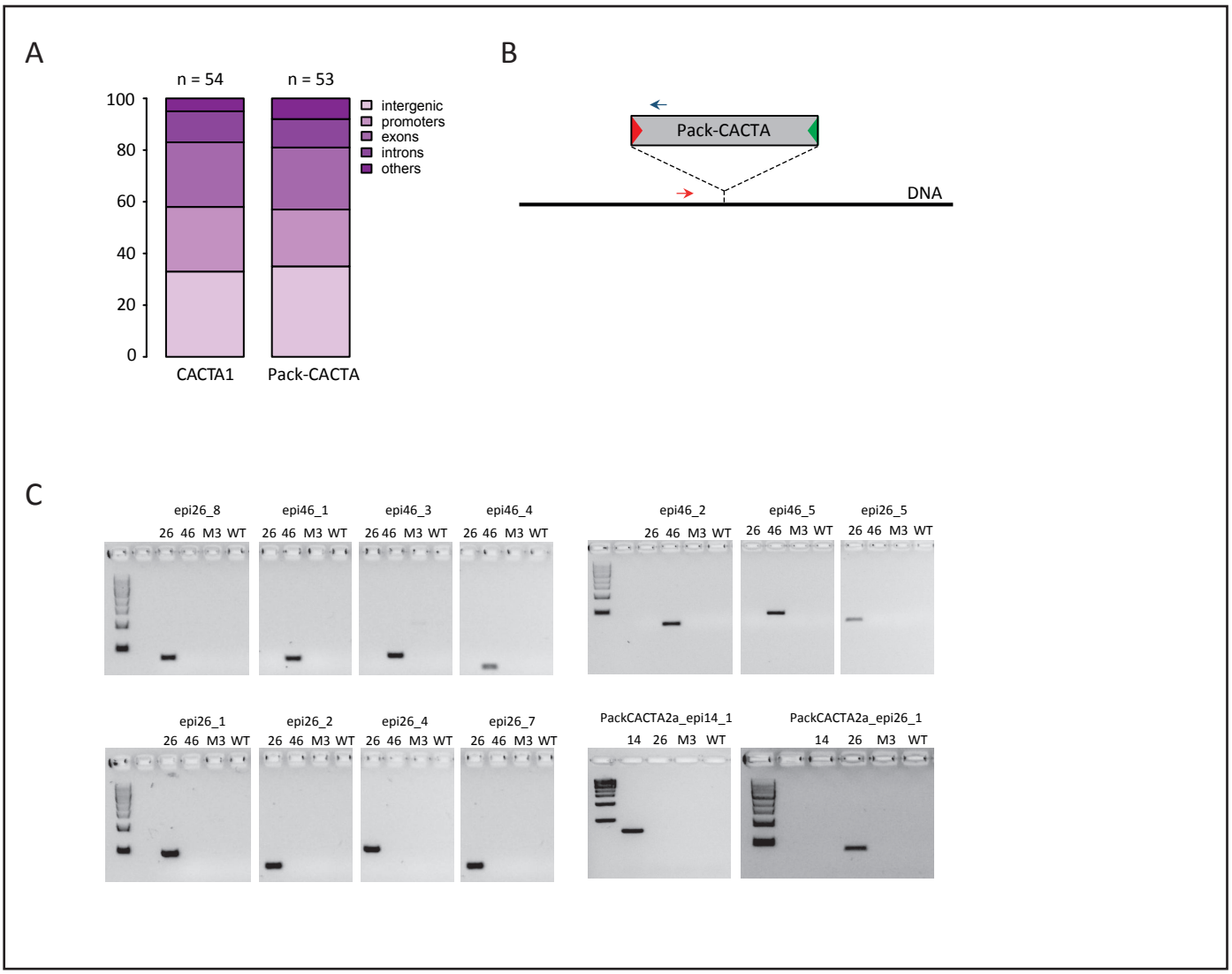

AT2G44470
epi26_2Δ
628 bp
mRNA 1
mRNA 2
AAAAA
AAAAA
AT3G01290
epi26_4Δ
264 bp
mRNA
AAAAA

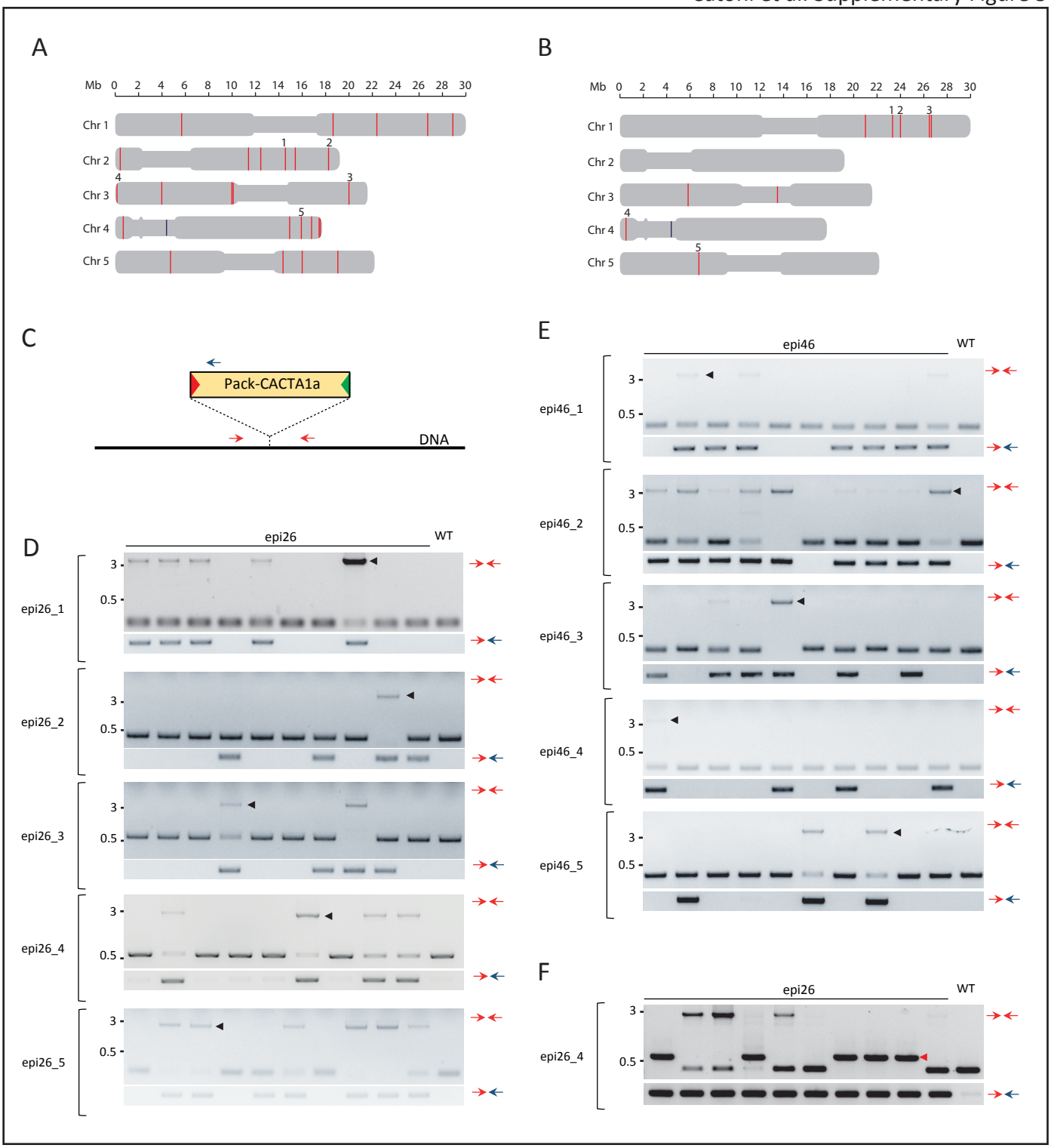
B

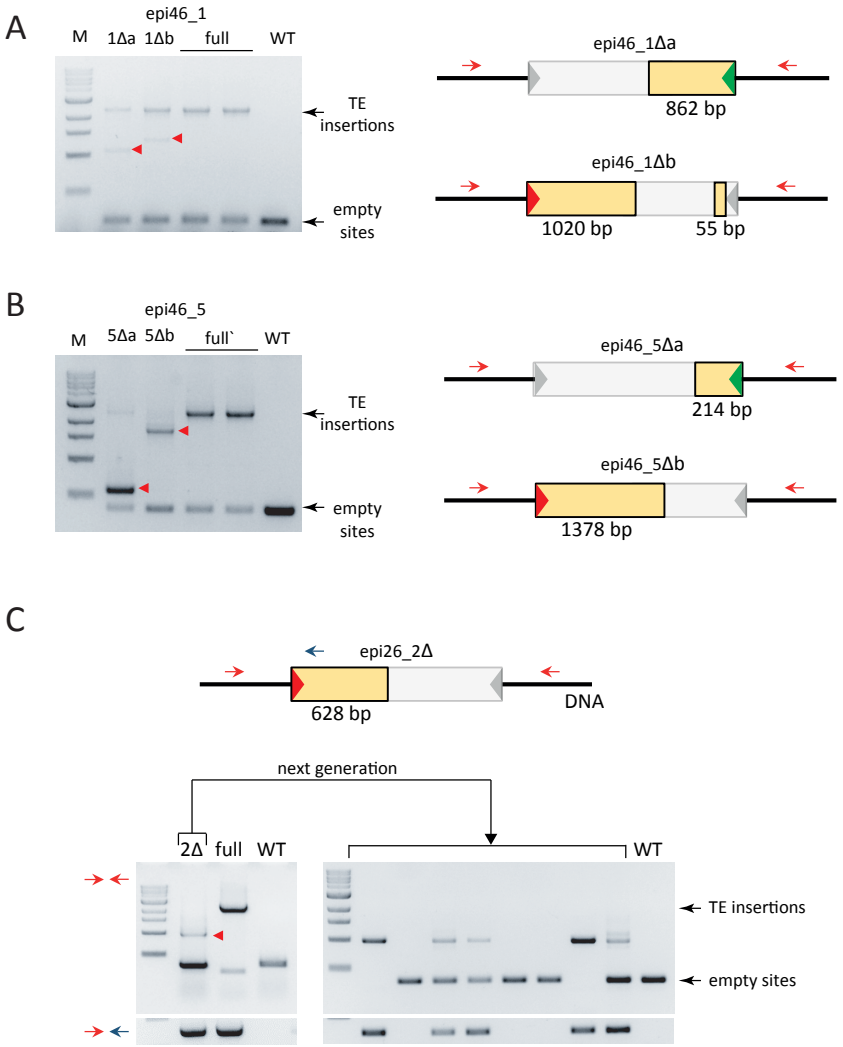

epi26
epi26_2Δ
HM
HET
empty
WT
0.5
1
0.5
UBQ
RT-
epi26
epi26_4Δ
HM
HET
empty
WT
1
0.5
0.5
UBQ
RT-


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

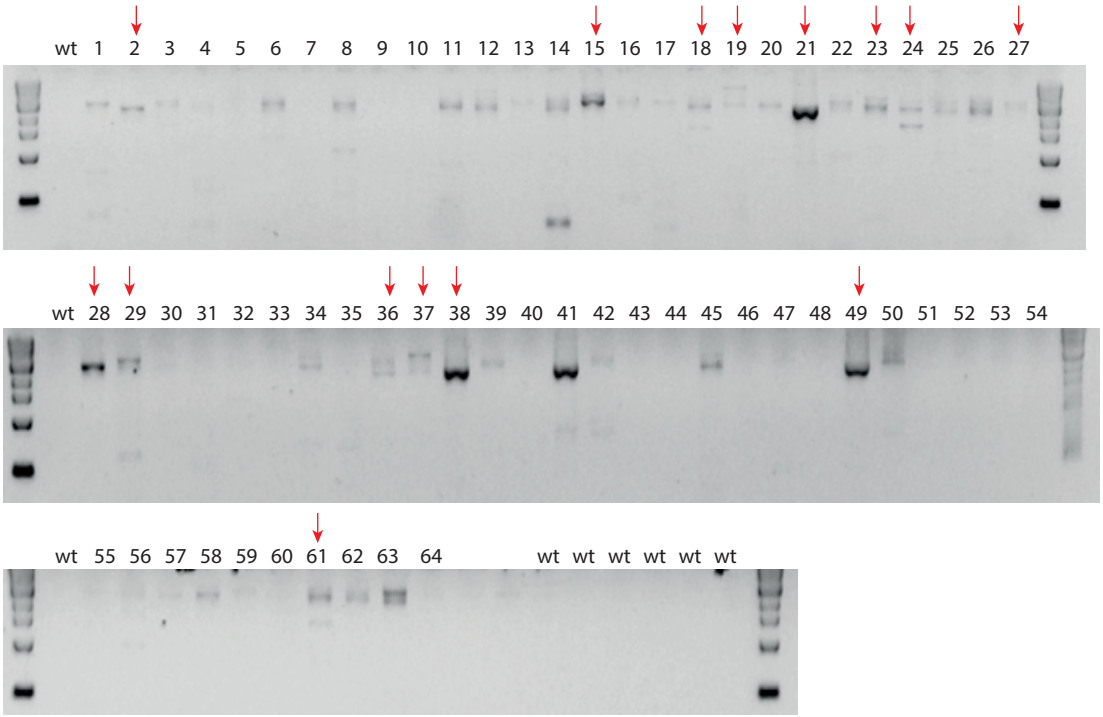

A
Chr 5
5,625
5,630
5,635
5,640
Kb
epiRILs
EVD active
01
02
03
04
05
epiRILs
EVD inactive
06
07
09
10
13
TE annnotation
EVADE' (AT5TE20395)
1Kb
B
Chr 2
4,900
4,905
4,910
Kb
epiRILs
CAC1 active
07
10
12
48
epiRILs
CAC1 inactive
01
02
04
05
epiRILs
CAC1 excision
36
44
45
21
TE annnotation
CACTA1 (AT2TE20205)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

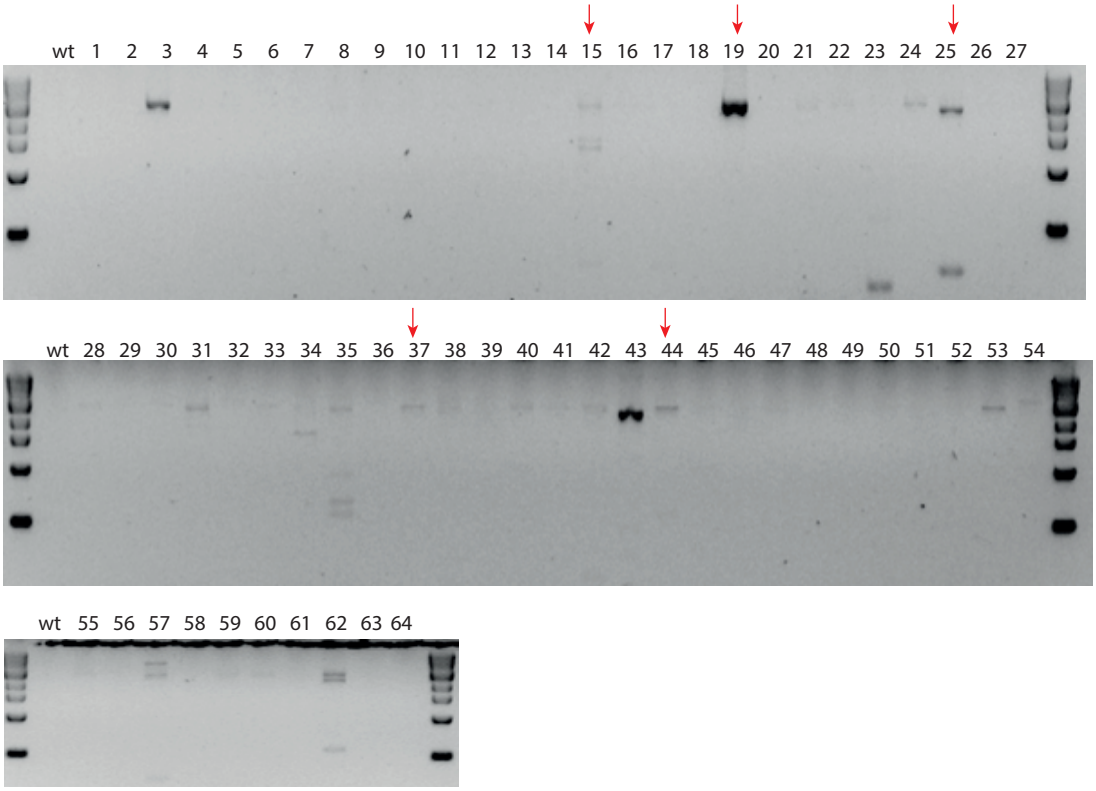

A
Mb 0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
B
C
Pack-CACTA1a
DNA
D
epi26
WT
epi26_1
epi26_2
epi26_3
epi26_4
epi26_5
3
0.5
E
epi46
epi46_1
epi46_2
epi46_3
epi46_4
epi46_5
F

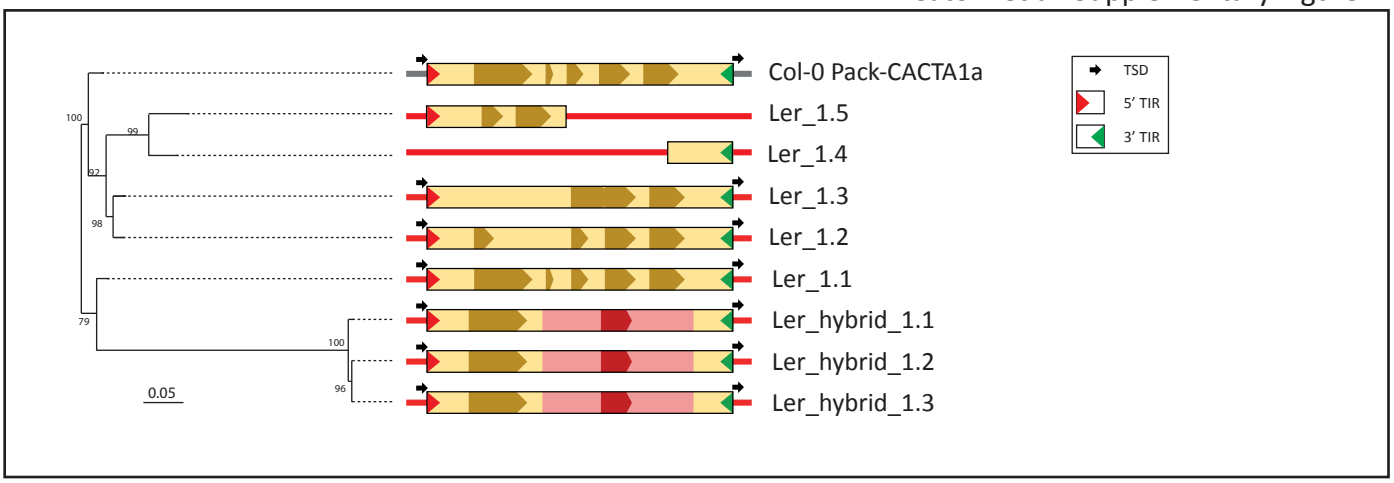

A
epi46_1
M
1Δa
1Δb
full
WT
TE insertions
empty sites
epi46_1Δa
862 bp
epi46_1Δb
1020 bp
55 bp
B
epi46_5
M
5Δa
5Δb
full`
WT
TE insertions
empty sites
epi46_5Δa
214 bp
epi46_5Δb
1378 bp
C
epi26_2Δ
628 bp
DNA
next generation
2Δ
full
WT
WT
TE insertions
empty sites

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