Molecular Plant
Letter to the Editor

Sequence-Independent Identification of Active LTR Retrotransposons in *Arabidopsis*

Dear Editor,

LTR retrotransposons (LTR-TEs) are mobile genetic elements that often comprise a large portion of the host genome and are particularly abundant in genomes of plants (Lisch, 2013). They transpose through an RNA transcript that is reverse transcribed to an extrachromosomal DNA (ecDNA) of the original element that can insert at a new genomic location (Wicker et al., 2007). Once activated, LTR-TEs can rapidly increase in copy number in the host genome, leading to mutations, genomic instability, and changes in the regulation of genes adjacent to new insertions (Lisch, 2013). The release of transcriptional silencing represents the first step of the LTR-TE life cycle, which does not predict subsequent steps leading to transposition. Therefore, it has been intrinsically difficult to reveal the movement of novel LTR-TEs in real time, and the identification of LTR-TEs capable of retrotransposition has mostly occurred during characterization of developmental defects induced by new integrations (Galindo-Gonzalez et al., 2017). Obviously, new retroelement insertions producing phenotypes make up a very small proportion of the total number of new integrations, and this approach has had some success only with highly active LTR-TEs with prevalent insertion in genic regions. Valuable alternatives for finding new active retrotransposons are genome-wide analyses revealing copy number increases at specific LTR-TEs loci or TE-associated structural variants (Tsukahara et al., 2009; Ewing, 2015; Lanciano et al., 2017). However, such approaches are expensive, labor, time intensive, and require a good quality reference genome. Furthermore, they can only reveal LTR-TEs already annotated or at least included in the reference genome. Here we present a method, which we named sequence-independent retrotransposon trapping (SIRT), which allows identification of LTR-TEs without previous knowledge of their DNA sequence.

We based our strategy on specific amplification and cloning of the linear double-stranded ecDNA produced by reverse transcription, i.e., the intermediate of the LTR-TE replication/transposition cycle, directly primed for re-integration into the genome (Figure 1A). To develop a method that detects ecDNA without previous knowledge of its sequence, we focused on the tRNA primer binding site (PBS) used for the priming step of reverse transcription. To date, all active LTR-TEs described in *Arabidopsis* use the initiator methionine tRNA (Met-iCAT) (Wicker et al., 2007). Consequently, their transcripts share a PBS sequence of 12 nucleotides complementary to the terminal nucleotides of the MET-iCAT tRNA. As a DNA fragment of 12 bp is too short for securing high specificity of PCR amplification with total genomic DNA, we lengthened the PBS-specific primers linking the PBS homology to complement the two terminal nucleotides of 5′ LTR, which mostly end in cytidine and adenosine. Since the distance between the end of the LTR and the PBS is variable,

we devised four primer sequences, which we termed Anchor0–3, corresponding to the LTR linked directly to the PBS or with one to three nucleotides in between, respectively (Figure 1B). A survey of the Arabidopsis TAIR10 reference genome showed that the four Anchor primers match 69% of all annotated LTR-TEs. Considering that LTR-TEs with no Met-iCAT matches are significantly smaller than LTR-TEs matching the Met-iCAT PBS (Figure 1C), they probably include many transposon remnants. Therefore, our primer combination is theoretically able to match the LTR sequences of 76% of the potentially active LTR-TEs in *Arabidopsis* (Figure 1D).

It is known that the LTR-TEs *ATCOPIA21* and *ATCOPIA93/EVADÉ* (*EVD*) become transpositionally active in hypomethylated conditions, such as in the *ddm1-2* mutant or in *met1*-derived epi-RILs, respectively (Mirouze et al., 2009; Tsukahara et al., 2009). In addition, the LTR-TE *COPIA78/ONSEN* was shown to produce ecDNA in wild-type and *nprd1–3* mutant plants as a consequence of heat stress (Ito et al., 2011). We investigated whether the SIRT approach could specifically trap the activity of *EVD*, *ATCOPIA21*, and *ONSEN* by their ecDNAs using the different Anchor primers. Using the Anchor0 primer, we identified PCR products corresponding in size to the LTRs of *EVD* with adaptor (469 bp) in epiRIL12 and *ddm1-2*, and to *COPIA21* LTR (184 bp) in *met1-1* (Figure 1E, Supplemental Figure 1A). Since *COPIA21* LTR ends in the sequence CACA, which is directly followed by the PBS, it can be amplified by both Anchor0 and Anchor2 (Figure 1E, Supplemental Figure 1A and 1C). Sequencing of bands confirmed the identity of the LTRs of *EVD* and *COPIA21* (Supplemental Figure 2). In heat-stressed samples of Col and *nrpd1-3*, bands detected only using the Anchor3 primer corresponded to the sizes of the LTRs of two *ONSEN* copies that differ in the length of their LTRs (503 and 526 bp). We sequenced three clones of this product and two aligned to *ONSEN* copy *At5g13205* and one to *ONSEN* copy *At1g11265*. As predicted, the sequences of all SIRT products included the entire length of the respective LTR linked to the adaptor sequence (Supplemental Figure 2). Using primers specific to *EVD*, *COPIA21*, and *ONSEN* in combination with the adaptor primer for the SIRT-ligated DNA, we further confirmed the presence of extrachromosomal DNA in the expected samples (Figure 1F). In the case of *COPIA21*, transposon display on two *met1-1* plants led to the identification of one new insertion in the *QUARTET 1* gene (*QRT1*, At5g55590) (Supplemental Figure 3).

Arabidopsis ecotype comparisons have shown that 80% of elements present in the Col-0 genome are absent in at least 1 of 80 sequenced accessions (Cao et al., 2011) thereby suggesting
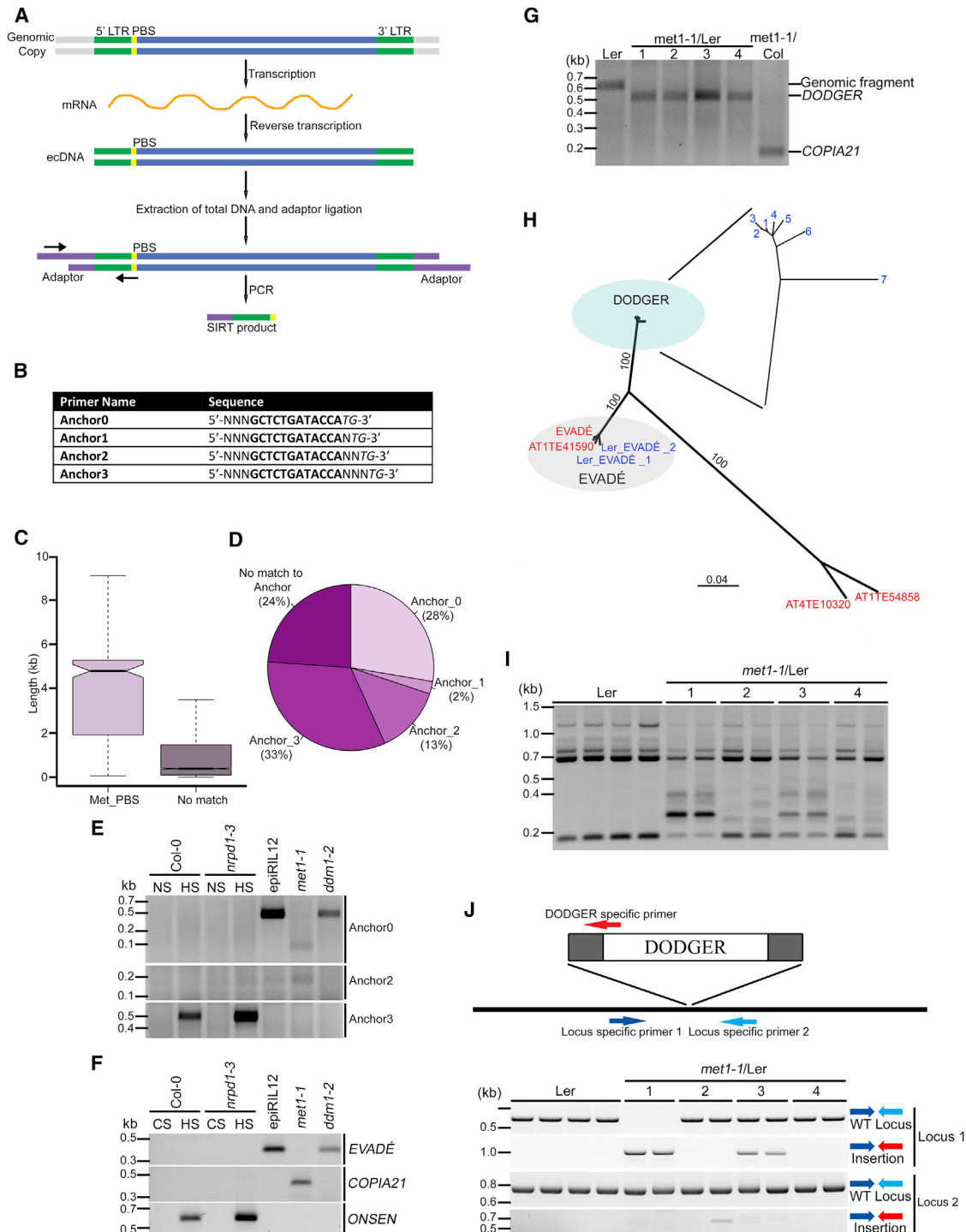
Figure 1. Sequence-Independent Retrotransposon Trapping Can Catch Active Elements.

(A) Scheme of the SIRT strategy. For amplification of linear ecDNA, total DNA is extracted and subjected to adaptor ligation. An adaptor-specific primer and an anchor primer designed on the conserved characteristics of the PBS region (arrows) were used. The expected product will be the length of the LTR plus the adaptor. LTRs are shown in green, the adaptors in purple, and the PBS in yellow.

(B) Anchor primer sequences. Bold, conserved 12-bp PBS sequence (Met-iCAT); italics, last 2 bp of the upstream 5′LTR, three random nucleotides added to the 5′ end to increase the annealing temperature.

(C) Size distributions of Col-0 annotated LTR-TEs either containing the Met-iCAT PBS (Met_PBS) or having no match to the Met-iCAT PBS (No match).

(D) Pie chart illustrating the matching potential of the Anchor primers (Anchor0 to Anchor3, Figure 1B) to the annotated LTR-TEs containing Met-iCAT PBS.

(E) PCR amplification with the adaptor-specific primer and individual anchor primers (marked on the right). Plant material used: Col-0 and *nrpd1-3* seedlings non-stressed (NS) and heat stressed (HS), seedlings (producing ecDNA of *ONSEN*), epiRIL12 (producing ecDNA of *EVADÉ*), *met1-1* (producing

(legend continued on next page)

Letter to the Editor

that other ecotypes might have novel transposable elements that are not present in the Col-0 reference. To test this hypothesis in a pilot experiment, we examined by SIRT a *Landsberg erecta* ecotype with the introgressed *met1-1* mutation (*met1-1*/Ler). Using the Anchor0 primer, we observed in *met1-1*/Ler plants a PCR product that was absent in Ler and *met1-1* (Col-0) plants and different in size to the *COPIA21* LTR (Figure 1G). We sequenced inserts of 448 bp, located between the end of the adaptor and the PBS primer, and found that they started with AATT or TATT. However, the remaining sequences were identical, suggesting that at least two different parental TEs contributed to the ecDNA (Supplemental Figure 4A). The insert sequences were blasted against the NCBI database. The fragment starting with TATT mapped with 100% identity to a resistance gene cluster of Ler, containing two retroelements (*La5-D* and *La5-G*) (Noel et al., 1999). The second best alignment was found in the Col-0 chromosome 3 where 405 bp of the SIRT insert mapped with 90% identity to a solo LTR with no other retroelement features, suggesting the absence of the Ler-inhabiting LTR-TE in the Col-0 background. We also sequenced an unspecific band in the Ler sample and found that it originated from inverted PBS sequences residing in the Ler genome. The absence of this band in the *met1-1*/Ler samples was most likely due to primer competition between genomic and ecDNA (Supplemental Figure 4B).

A BLAST search to the *de novo* assembled Ler-0 genome from long reads of the PacBio (http://www.pacb.com/) revealed the presence of five retrotransposons with 100% identity to one of the two LTRs recovered by SIRT and a further two with higher than 97% identity. Furthermore, we found two further closely related elements with 76.2% and 75% identity to the SIRT inserts. The LTR sequences of these two last elements shared 94.7% identity with the LTR of the previously characterized retroelement *COPIA93/EVD*, which is active in epiRILs constructed in the Col-0 background (Mirouze et al., 2009). The alignments of all mentioned retrotransposon sequences revealed a clear distance between *EVD* and the new Ler-specific elements (Figure 1H), which we named *DODGER*. The seven elements of the *DODGER* family are all young retrotransposons, with a maximal one mismatch difference between the two LTRs (Supplemental Figure 5). We amplified and sequenced the adapter ligated ecDNA using primers located internally in the *DODGER* sequences, and using informative SNPs, we restricted the elements able to produce ecDNA to *DODGER* 1, 2, and 3.

Finally, to determine whether *DODGER* is transposition competent, we carried out transposon display on two independent pools consisting of five progeny plants from each of the four parental plants initially tested by SIRT. Multiple new bands appeared in *met1-1*/Ler that were absent in the Ler wild-type, suggesting that new integrations of *DODGER* occurred (Figure 1I). Sequences obtained from two of these bands were blasted against the Ler PacBio genome to identify the insertion loci. We confirmed both integration events using primer pairs designed from the DNA flanking the *DODGER* new insertions and a *DODGER*-specific primer (Figure 1J). We amplified and cloned both LTRs from each of the two newly integrated copies and found that both *DODGER* new insertions now have 100% identity between the LTRs and a 5-bp target site duplication characteristic of new COPIA insertions (Wicker et al., 2007) (Supplemental Figures 6 and 7). Furthermore, using informative SNPs, we identified the new insertions as *DODGER*_1 (insertion 2) and *DODGER*_3 (insertion 1) (Supplemental Figure 7).

We propose SIRT is the optimal tool to identify novel active elements in a cost-effective manner in non-sequenced background in species with a relatively low content of LTR-TEs. Contrary to next-generation sequencing approaches, SIRT does not require TE annotation or an assembled genome. In comparison with the recently developed mobilome sequencing (Lanciano et al., 2017), which detects circular by-products of active LTR retrotransposons through random primer amplification of all circular DNAs present in a given sample and genome alignments, SIRT directly traps the 5' LTR of linear extrachromosomal DNA of LTR retrotransposons through PBS priming and ligation of the adapter. The sequence of the trapped LTR and thus later of unknown transposons can be revealed by direct Sanger sequencing of the SIRT product(s). Therefore, these two methods significantly differ, but at the same time they are highly complementary. Importantly, the anchor primers used here have been designed for *Arabidopsis*, in consequence the application of SIRT to other organisms may require certain primer optimizations. Nonetheless, the use of the universal prevalence of PBS sequences matching the methionine tRNA, as described here, seems to be the best option for an initial approach to LTR-TE detection in an unknown genome.

SUPPLEMENTAL INFORMATION
Supplemental Information is available at *Molecular Plant Online*.

---

ecDNA of *COPIA21*), and *ddm1-2* (producing ecDNA of *EVADÉ*). Complete gels are located in Supplemental Figure 1, including the result of the Anchor 1 primer amplification, which did not amplify any products.

(F) PCR confirmation of the identity of the ecDNA using transposon-specific primers with the adaptor-specific primer on the adaptor-ligated DNA. PCR products of the expected sizes for each element (marked on the right) were obtained. Col-0 and *nrpd1-3* seedlings non-stressed (NS) and heat stressed (HS), seedlings, epiRIL12, *met1-1*, and *ddm1-2*.

(G) SIRT analysis of the DNA from Ler of *Arabidopsis*. DNA was extracted from 2-week-old seedlings of Ler wild-type and Ler with the introgressed *met1-1* mutation (*met1-1*/Ler). PCR was carried out using the adaptor-specific primer and Anchor0. Amplified fragments were subsequently sequenced and their identities are indicated on the right (sequences in Supplemental Figure 4).

(H) Phylogenetic tree consisting of seven *DODGER*, two Col *EVADÉ*, and two Ler *EVADÉ* elements demonstrating that *DODGER* is a family distinct from *EVADÉ*. Ler-derived sequences are in blue and Col-0-derived sequences in red.

(I) Transposon display analysis on genomic DNA digested by DraI. Adaptor-specific and *DODGER*-specific primers were used for detecting new *DODGER* insertions in the four *met1-1*/Ler lines.

(J) Validation of two new *DODGER* insertions in the *met1-1*/Ler background. Primer locations for PCR are indicated above the gel image and their uses to the right (chromosomal locus-specific primers are indicated in blue and *DODGER*-specific primers in red). The insertion in locus 1 is homozygous and heterozygous in lines 1 and 3, respectively. The insertion in locus 2 is only present in one pool of plants from line 2.

Molecular Plant Letter to the Editor

Jayne Griffiths[1], Marco Catoni[1],
Mayumi Iwasaki[1,2] and Jerzy Paszkowski[1],*

[1]The Sainsbury Laboratory, University of Cambridge, Cambridge, UK
[2]Present Address: University of Geneva, Geneva, Switzerland
*Correspondence: Jerzy Paszkowski (jerzy.paszkowski@slcu.cam.ac.uk)
https://doi.org/10.1016/j.molp.2017.10.012

REFERENCES

Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., et al. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat. Genet. **43**:956–963.

Ewing, A.D. (2015). Transposable element detection from whole genome sequence data. Mob. DNA **6**:24.

Galindo-Gonzalez, L., Mhiri, C., Deyholos, M.K., and Grandbastien, M.A. (2017). LTR-retrotransposons in plants: engines of evolution. Gene **626**:14–25.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature **472**:115–119.

Lanciano, S., Carpentier, M.C., Llauro, C., Jobet, E., Robakowska-Hyzorek, D., Lasserre, E., Ghesquiere, A., Panaud, O., and Mirouze, M. (2017). Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. PLoS Genet. **13**:e1006630.

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

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

Noel, L., Moores, T.L., van Der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E., and Jones, J.D. (1999). Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of *Arabidopsis*. Plant Cell **11**:2099–2112.

Tsukahara, S., Kobayashi, A., Kawabe, A., Mathieu, O., Miura, A., and Kakutani, T. (2009). Bursts of retrotransposition reproduced in *Arabidopsis*. Nature **461**:423–426.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. **8**:973–982.