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 a 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 insertions (Galindo-Gonzalez et al., 2017). Obviously, new retroelements 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 (Tsuchahara 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/EVADE(EVD) become transpositionally active in hypomethylated conditions, such as in the ddm1-2 mutant or in met1-derived epiRILs, respectively (Mirouze et al., 2009; Tsuchahara 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 nprd1–3, 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).

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

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.
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 EVADE), met1-1 (producing...
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 TE contributions 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 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 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 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.
FUNDING
This work was supported by the European Research Council (EVOBREED; 322621); and Gatsby Fellowship (AT3273/GLE).

AUTHOR CONTRIBUTIONS
Conceptualization, J.P. and M.I.; Methodology, J.G. and M.C.; Investigation, J.G. and M.C.; Writing – Original Draft, J.G., M.C., and J.P.; Writing – Review & Editing, J.G., M.C., and J.P.; Funding Acquisition, J.P.; Supervision, J.P.

ACKNOWLEDGMENTS
We are grateful to Raymond Wightman (SLCU Core Microscopy Facility) for assisting with the SEM microscopy and Eric Richards for providing the met1-1/Ler seeds. No conflict of interest declared.

Received: August 18, 2017
Revised: September 28, 2017
Accepted: October 19, 2017
Published: October 26, 2017

Jayne Griffiths¹, Marco Catoni¹, Mayumi Iwasaki¹,² and Jerzy Paszkowski¹,∗
¹The Sainsbury Laboratory, University of Cambridge, Cambridge, UK
²Present 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., Mirozu, M., Vaillant, I., and Paszkowski, J. (2011). An sRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472:115–119.

Lanciano, S., Carpentier, M.C., Llauro, C., Jobet, E., Robakowska-Hyzerk, D., Lasserre, E., Ghesquiere, A., Panaud, O., and Mirozu, 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.

Mirozu, 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.