

The Arabidopsis genus

An emerging model to elucidate the molecular basis of interspecific differences in transposable element activity

Juliette de Meaux^{1,*} and Ales Pecinka²

¹Institute for Evolution and Biodiversity; University of Münster; Münster, Germany; ²Department of Plant Breeding and Genetics; Max Planck Institute for Plant Breeding Research; Cologne, Germany

Keywords: transposable element, transcription, genome size, epigenetics, cis-regulation

Arabidopsis thaliana is a model plant species and its molecular dissection has greatly contributed to our understanding of the systems preventing genome invasion by transposable elements (TE). Recent advances suggest that *A. thaliana* may be more efficient than its congener *A. lyrata* at controlling TE expression and proliferation. The comparative analysis of TE transcription in *A. thaliana* and *A. lyrata*, which differ by 40% in genome size, may help understand how silencing mechanisms contribute to the evolution of transposition rate, an important factor controlling genome size variation in plants and animals.

The number of transposable elements (TEs) is known to vary greatly among plant genomes and evolves as the net outcome of three major factors: transposition activity, TE removal and efficiency of purifying selection as determined by population genetics processes.^{1,3} In the genus *Arabidopsis*, the two species *A. lyrata* and *A. thaliana* differ by ~40% in genome size.^{4,5} As much as 56% of the *A. lyrata* genome sequence that cannot be aligned to *A. thaliana* encodes TEs or simple repeats.⁵ Differences in TE content therefore explain a significant fraction of genome size differences. Out-crossing species are expected to be more efficient in removing deleterious mutations than inbred species.¹ As a consequence, new transposition variants with deleterious effects on plant fitness should segregate at lower frequency in the species that can best purify deleterious mutations, i.e., in the outbred *A. lyrata*. However, transposition variants are not segregating at lower frequency in *A. lyrata*.^{6,7} Therefore, population dynamics of deleterious mutations do not influence significantly the difference shown by *A. thaliana* and *A. lyrata* in TE content.²

Instead, evidence is accumulating that the two species differ in the transcriptional control of TEs. In the model plant species *A. thaliana*, this control is increasingly well understood.⁸⁻¹¹ Silencing of TEs is mediated by small interfering RNAs (siRNAs) that guide the deposition of DNA and histone methylation marks on homologous DNA stretches.¹²⁻¹⁴ These marks then act in *cis*- to repress expression. In *A. thaliana*, most TEs reside in

pericentromeric chromosomal regions that are transcriptionally inactive.^{15,16} Under stressful conditions, silencing can be transiently suppressed, making room for potential transposition.^{17,18,19} Knocking-out distinct components of the transcriptional gene silencing system can also activate specific subsets of TEs.^{15,20} The analysis of small RNAs expressed in both species showed that *A. thaliana* appears to be able to repress TE transcription more efficiently than *A. lyrata*. The specificity of siRNAs for their target TEs is strongly associated with the efficacy of TE silencing in both *A. lyrata* and *A. thaliana*, but *A. thaliana* shows a 3-fold greater proportion of siRNAs mapping uniquely to a single TE copy.²¹ This finding agrees with the observation that a greater number of TEs are expressed in *A. lyrata*.²¹ In *A. thaliana*, genes located less than 2.5 kb away from TEs tend to be less expressed, suggesting that silencing TEs can entail negative consequences for plant fitness if protein-coding genes in the vicinity of TE insertions are partially silenced and cannot be expressed properly.²² This cost seems weaker in the congeneric species *A. lyrata* where TEs are more abundant and tend to be closer to expressed genes. Indeed, in this species, TE insertions are associated with a reduced expression of their closest protein-coding neighbor only if they are located less than 1.5 kb away.²¹ Interestingly, protein-coding genes showing greater expression of the *A. lyrata* allele in F1 interspecific hybrids confirm that the silencing effect of TE proximity is greater in *A. thaliana*.²³

Using a novel approach to monitor genome-wide variation in TE *cis*-regulation, He et al. in the de Meaux laboratory have demonstrated major interspecific differences in the *cis*-regulation of TE silencing.²⁴ In F1 interspecific hybrids, monitoring of allele-specific expression at 1535 loci annotated as TEs in *A. thaliana* showed that as many as 47% of TE loci display allele-specific expression and thus differ in their *cis*-regulation. Interestingly, almost all differentially expressed TEs expressed the *A. lyrata* allele more than the *A. thaliana* allele. Allele-specific expression of TEs is not observed between *A. thaliana* accessions.²⁵ He, de Meaux and collaborators²⁴ confirmed on a sub-sample of 18 loci that the upregulation of *A. lyrata* alleles in F1 interspecific hybrids reflects parental differences in TE expression, and is not simply a result of TE upregulation induced by “genomic shock,” a phenomenon often observed in interspecific hybrids or synthetic *Arabidopsis* allopolyploids.²⁶ A previous analysis based on RNA-sequencing data reported that 8% of *A. lyrata* TE loci are

*Correspondence to: Juliette de Meaux;
Email: juliette.de.meaux@uni-muenster.de
Submitted: 05/29/12; Accepted: 06/12/12
<http://dx.doi.org/10.4161/mge.21111>

expressed,²¹ but the study of He et al.²⁴ might be more sensitive as the approach is not affected by coverage issues that prevent the detection of lowly expressed genes. Importantly, this study further supports the hypothesis proposed by Hollister et al.²¹ that interspecific differences in TE regulation depend on *cis*-acting marks deposited by the epigenetic machinery. For ten of 11 TEs examined, H3K9me2 methylation, a silencing histone mark, was detected in *A. thaliana* but not in *A. lyrata*.²⁴ In addition, differentially regulated TEs were significantly enriched among TEs controlled by the DNA methyl-transferase MET1, suggesting that allele-specific expression of TEs depends on CG methylation.^{24,27} Since *A. lyrata* TE alleles are systematically upregulated, it is possible that unknown changes in the epigenetic machinery of the two species create genome-wide differences in epigenetic marks, which are maintained in interspecific F1 hybrids¹¹ and act in *cis*- to preferentially silence the *A. thaliana* allele. We suspect that interspecific differences in the epigenetic machinery are mostly quantitative: *A. lyrata* MET1 shares 90% amino acid identity with its *A. thaliana* ortholog (Pecinka A, unpublished data) and appears functional. In addition, *A. thaliana* and *A. lyrata* seem to have the same repertoire of epigenetic marks,^{24,28} although the *A. lyrata* epigenome has not yet been characterized. Further experiments are warranted to confirm this hypothesis. In fact, sequence differences in TE regulatory regions probably also contribute to methylation differences between TE alleles. The endosperm expressed-gene *FWA*, for example, shows interspecific variation in patterns of epigenetic silencing likely driven by the presence of repeat sequences in the promoter²⁸ and differential TE regulation associates with nucleotide differences in TE upstream regions.²⁴

TE transcriptional upregulation is sometimes associated with increased transposition rates.² Do the *cis*-regulatory differences observed for Arabidopsis TEs²⁴ reflect a difference in transposition activity? The genomic sequence of *A. lyrata* revealed that recently inserted TEs were more abundant in *A. lyrata* than in *A. thaliana*.^{5,16,29} Therefore, the genome size difference between the two species is not only due to a massive removal of TEs in *A. thaliana*: TE amplification in *A. lyrata* also plays a role. He et al.²⁴ found that TEs showing differences in *cis*-regulation do not show a comparatively higher number of copies. Using published age estimates,⁵ they observed that LTR retrotransposons differentially expressed in F1 hybrids are among the youngest TEs in the *A. thaliana* genome. This suggests that *A. thaliana* silences recent TE copies more effectively, although older copies are similarly regulated in both species, a finding in agreement with the interspecific differences in uniquely mapping siRNAs mentioned above.²¹ Nonetheless, TE age was estimated based on the divergence of the terminal repeats in LTR retrotransposons, which are identical at the time of insertion.³⁰ We observed that age estimates are not congruent between species. Orthologous LTR retrotransposons, defined as elements flanked by orthologous neighbor

genes, were estimated to be older in *A. lyrata*. Since orthologous TE insertions should be the same age, this questions the accuracy of LTR age estimation. Population parameters, which differ widely between outcrossing *A. lyrata* and selfing *A. thaliana*,³¹ probably influence the rate of LTR divergence, making interspecific TE age comparisons difficult. In addition, age estimates would be flawed if LTRs undergo gene conversion.³⁰ To our knowledge, the role of this potential confounding factor has not been considered in these two species. The age distribution of LTR retrotransposons in the two species therefore needs to be re-examined in greater depth. The existence of inter-specific differences in transcriptional silencing of TEs is now clear but whether they have caused the widely different TE contents of the genomes of *A. lyrata* and *A. thaliana* remains to be demonstrated.

Interestingly, not only the host silencing system is evolving in the Arabidopsis genus: the elements themselves have evolved new functions that may also contribute to interspecific differences in TE proliferation. A groundbreaking study has demonstrated that orthologous families of TEs can evolve different transposition site preferences. Tsukahara et al.²⁹ introduced *Tall*, an *A. lyrata* member of the *COPIA93/Evade* TE family,^{15,16} into the *A. thaliana* genome and monitored its activity. In *A. lyrata*, this TE recently proliferated in centromeric regions. In contrast, *COPIA93* members in *A. thaliana* are found in low copy numbers and insert in chromosome arms. When introduced into *A. thaliana*, *Tall* is not efficiently silenced, and proliferates by integrating specifically in the centromeres. The rate of *Tall* transposition is magnified in a *ddm1* mutant background, suggesting that *Tall* does not entirely escape the *A. thaliana* defense system.²⁹ The release of *COPIA93* silencing in *A. thaliana ddm1* mutants leads to new integrations in chromosome arms but never in centromeres. Preferential insertion into centromeric regions seems to be an ancestral property of *Tall* and thus *COPIA93* shows a modified insertion site preference in the *A. thaliana* lineage. This experiment provides an admirable demonstration that TEs have evolved new insertion abilities since the separation of *A. lyrata* and *A. thaliana*.

The rich molecular and genomic toolbox available in *A. thaliana* and the availability of the *A. lyrata* genome should facilitate the identification of molecular factors controlling both host and invader variation. Interspecific variation in the Arabidopsis genus therefore promises to bring novel insights into the mechanisms controlling the evolution of both “selfish DNA” and its control by the host’s genomic system. It also promises to shed light on a long-standing question: whether selfish DNA and the host defense system coevolve.

Acknowledgments

We thank L. Smith (MPI Tübingen) for feedback on the manuscript. This work was funded by the Deutsche Forschungsgemeinschaft within the realm of the priority program SPP1529 “Adaptomics.”

References

1. Tenaillon MI, Hollister JD, Gaut BS. A triptych of the evolution of plant transposable elements. *Trends Plant Sci* 2010; 15:471-8; PMID:20541961; <http://dx.doi.org/10.1016/j.tplants.2010.05.003>.
2. Ågren JA, Wright SI. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chromosome Res* 2011; 19:777-86; PMID:21850458; <http://dx.doi.org/10.1007/s10577-011-9229-0>.
3. Le Rouzic A, Boutin TS, Capy P. Long-term evolution of transposable elements. *Proc Natl Acad Sci U S A* 2007; 104:19375-80; PMID:18040048; <http://dx.doi.org/10.1073/pnas.0705238104>.
4. Oyama RK, et al. The shrunken genome of *Arabidopsis thaliana*. *Plant Syst Evol* 2008; 273:257-71; <http://dx.doi.org/10.1007/s00606-008-0017-z>.
5. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* 2011; 43:476-81; PMID:21478890; <http://dx.doi.org/10.1038/ng.807>.
6. Lockton S, Ross-Ibarra J, Gaut BS. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A* 2008; 105:13965-70; PMID:18772373; <http://dx.doi.org/10.1073/pnas.0804671105>.
7. Lockton S, Gaut BS. The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evol Biol* 2010; 10:10; PMID:20067644; <http://dx.doi.org/10.1186/1471-2148-10-10>.
8. Zilberman D, Henikoff S. Silencing of transposons in plant genomes: kick them when they're down. *Genome Biol* 2004; 5:249; PMID:15575975; <http://dx.doi.org/10.1186/gb-2004-5-12-249>.
9. Zilberman D, Henikoff S. Epigenetic inheritance in *Arabidopsis*: selective silence. *Curr Opin Genet Dev* 2005; 15:557-62; PMID:16085410; <http://dx.doi.org/10.1016/j.gde.2005.07.002>.
10. Weil C, Martienssen R. Epigenetic interactions between transposons and genes: lessons from plants. *Curr Opin Genet Dev* 2008; 18:188-92; PMID:18339541; <http://dx.doi.org/10.1016/j.gde.2008.01.015>.
11. Feng S, Jacobsen SE, Reik W. Epigenetic reprogramming in plant and animal development. *Science* 2010; 330:622-7; PMID:21030646; <http://dx.doi.org/10.1126/science.1190614>.
12. Daxinger L, Kanno T, Bucher E, van der Winden J, Naumann U, Matzke AJ, et al. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J* 2009; 28:48-57; PMID:19078964; <http://dx.doi.org/10.1038/emboj.2008.260>.
13. Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, et al. The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* 2010; 22:321-34; PMID:20173091; <http://dx.doi.org/10.1105/tpc.109.072199>.
14. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 2010; 11:204-20; PMID:20142834; <http://dx.doi.org/10.1038/nrg2719>.
15. Mirouze M, Reinders J, Bucher E, Nishimura T, Schneberger K, Ossowski S, et al. Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* 2009; 461:427-30; PMID:19734882; <http://dx.doi.org/10.1038/nature08328>.
16. Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* 2009; 461:423-6; PMID:19734880; <http://dx.doi.org/10.1038/nature08351>.
17. Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*. *Plant Cell* 2010; 22:3118-29; PMID:20876829; <http://dx.doi.org/10.1105/tpc.110.078493>.
18. Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I. Stress-induced activation of heterochromatic transcription. *PLoS Genet* 2010; 6:e1001175; PMID:21060865; <http://dx.doi.org/10.1371/journal.pgen.1001175>.
19. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 2011; 472:115-9; PMID:21399627; <http://dx.doi.org/10.1038/nature09861>.
20. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* 2001; 411:212-4; PMID:11346800; <http://dx.doi.org/10.1038/35075612>.
21. Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A* 2011; 108:2322-7; PMID:21252301; <http://dx.doi.org/10.1073/pnas.1018222108>.
22. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 2009; 19:1419-28; PMID:19478138; <http://dx.doi.org/10.1101/gr.091678.109>.
23. He F, Zhang X, Hu J, Turck F, Dong X, Goebel U, et al. Genome-wide analysis of cis-regulatory divergence between species in the *Arabidopsis* genus. *Mol Biol Evol* 2012; •••; <http://dx.doi.org/10.1093/molbev/mss146>; PMID:22641789.
24. He F, Zhang X, Hu JY, Turck F, Dong X, Goebel U, et al. Widespread interspecific divergence in cis-regulation of transposable elements in the *Arabidopsis* genus. *Mol Biol Evol* 2012; 29:1081-91; PMID:22086904; <http://dx.doi.org/10.1093/molbev/mst281>.
25. Zhang X, Borevitz JO. Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics* 2009; 182:943-54; PMID:19474198; <http://dx.doi.org/10.1534/genetics.109.103499>.
26. Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, et al. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J* 2005; 41:221-30; PMID:15634199; <http://dx.doi.org/10.1111/j.1365-313X.2004.02297.x>.
27. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 2008; 133:523-36; PMID:18423832; <http://dx.doi.org/10.1016/j.cell.2008.03.029>.
28. Fujimoto R, Kinoshita Y, Kawabe A, Kinoshita T, Takashima K, Nordborg M, et al. Evolution and control of imprinted FWA genes in the genus *Arabidopsis*. *PLoS Genet* 2008; 4:e1000048; PMID:18389059; <http://dx.doi.org/10.1371/journal.pgen.1000048>.
29. Tsukahara S, Kawabe A, Kobayashi A, Ito T, Aizu T, Shin-i T, et al. Centromere-targeted de novo integrations of an LTR retrotransposon of *Arabidopsis lyrata*. *Genes Dev* 2012; 26:705-13; PMID:22431508; <http://dx.doi.org/10.1101/gad.183871.111>.
30. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. The paleontology of intergene retrotransposons of maize. *Nat Genet* 1998; 20:43-5; PMID:9731528; <http://dx.doi.org/10.1038/1695>.
31. Wright SI, Le QH, Schoen DJ, Bureau TE. Population dynamics of an Ac-like transposable element in self- and cross-pollinating *Arabidopsis*. *Genetics* 2001; 158:1279-88; PMID:11454774.