

Review Article

Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs

Yichun Qiu and  Claudia Köhler

Department of Plant Biology, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Uppsala BioCenter, Almas Allé 5, Uppsala, Sweden

Correspondence: Claudia Köhler (Claudia.kohler@slu.se)

Transposable elements (TEs) constitute major fractions of plant genomes. Their potential to be mobile provides them with the capacity to cause major genome rearrangements. Those effects are potentially deleterious and enforced the evolution of epigenetic suppressive mechanisms controlling TE activity. However, beyond their deleterious effects, TE insertions can be neutral or even advantageous for the host, leading to long-term retention of TEs in the host genome. Indeed, TEs are increasingly recognized as major drivers of evolutionary novelties by regulating the expression of nearby genes. TEs frequently contain binding motifs for transcription factors and capture binding motifs during transposition, which they spread through the genome by transposition. Thus, TEs drive the evolution and diversification of gene regulatory networks by recruiting lineage-specific targets under the regulatory control of specific transcription factors. This process can explain the rapid and repeated evolution of developmental novelties, such as C_4 photosynthesis and a wide spectrum of stress responses in plants. It also underpins the convergent evolution of embryo nourishing tissues, the placenta in mammals and the endosperm in flowering plants. Furthermore, the gene regulatory network underlying flower development has also been largely reshaped by TE-mediated recruitment of regulatory elements; some of them being preserved across long evolutionary timescales. In this review, we highlight the potential role of TEs as evolutionary toolkits in plants by showcasing examples of TE-mediated evolutionary novelties.

Introduction

Transposable elements (TEs) are parasitic self-replicating elements. They contribute to large fractions of host genomes as they continuously amplify and disperse themselves into different loci. TEs also have been reshaping genomes by causing major genome rearrangements [1–4].

There are two major classes of TEs that differ in their transposition mechanism [1,5]. Class I TEs, or retrotransposons, often have long terminal repeats (LTRs) at each end and transpose via an RNA intermediate that is reverse-transcribed and then integrates back into the genome. Non-LTR retrotransposons include long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Class II TEs, or DNA transposons, are often flanked by terminal inverted repeats (TIRs) and rely on DNA intermediates for transposition by employing a cut and paste mechanism. Helitrons are class II TEs that do not contain TIRs and transpose by a rolling circle replication mechanism by inserting a single-stranded circular DNA intermediate into a new genomic locus [6–8]. Both classes of TEs include autonomous and non-autonomous elements. While autonomous elements encode the complete set of enzymes required for transposition, non-autonomous elements rely on the enzymes of autonomous TEs to transpose [9]. Based on defined criteria, class I and II TEs can be subdivided into multiple superfamilies [10].

Received: 22 April 2020

Revised: 1 June 2020

Accepted: 3 June 2020

Version of Record published:
23 June 2020

TEs wire transcriptional networks

TE remobilization can cause multiple effects; the most direct effects are mutations due to TE insertions into genes or regulatory elements and chromosome instability [2,9]. These potential deleterious effects caused by TE remobilization have enforced the evolution of multiple epigenetic mechanisms to recognize and silence TEs and TE-derived repetitive sequences [11–15]. Nevertheless, TE insertions can be neutral or even advantageous for the host, leading to long-term retention of TEs in the host genome. The latter case, referred to as TE exaptation, or TE domestication, is a process where TEs become part of the regulatory and/or coding region of host genes and contribute to diverse phenotypes and important functions [16–18].

Insertions of TEs can have a strong regulatory impact on the expression of adjacent genes. This can be mediated by multiple mechanisms, including the alteration of chromatin structure, disruption of *cis*-regulatory elements, or providing novel regulatory elements [9,19–23]. In this review, we will focus on the evolutionary contribution of TEs to generate new regulatory networks and thus developmental novelties by distributing regulatory elements in plants. For more details on TE-mediated gene regulation and other aspects of TE domestication we refer the reader to recent reviews covering these topics [19,24].

TEs have long been proposed as attractive vehicles accounting for the evolution of eukaryotic gene regulatory networks [25,26]. This idea has gained support from studies in animals and plants, linking the expansion of certain TE families with the dispersal of regulatory elements required for specific developmental programs [27–30]. One striking example in animals is the evolution of the mammalian placenta, the nourishing tissue supporting embryo growth, which has been linked to the activities of endogenous retroviruses, or LTR retrotransposons; many families of which have been domesticated to generate novel regulatory genes or to convey regulatory elements [31,32]. Out of many placenta-related TEs, two distinct groups of retroviral LTR retrotransposons, *THE1B* and *RLTR13D5*, have contributed to the lineage-specific dispersal of hundreds of placenta-specific regulatory elements in anthropoid primates and mice, respectively, which likely contributed to the rapid morphological diversification of the mammalian placenta (Figure 1) [33,34].

Interestingly, the convergent evolution of the endosperm, the nourishing tissue in flowering plants, was also promoted by TE transpositions (Figure 1). The endosperm-specific type I MADS-box transcription factor PHERES1 (PHE1) in *Arabidopsis thaliana* binds to two kinds of motifs, one resembling canonical MADS-box associated CArG motifs, while the other representing partially modified CArG motifs, which are both enriched in helitron elements [35]. Helitron sequences in the *A. thaliana* genome have a higher density of CArG motif sequences than other TE sequences and the rest of the genome, consistent with their proposed role as vehicles of such binding motifs. Many targeted genes with helitron-derived PHE1 binding motifs have high expression in the endosperm, reflecting endosperm-specific regulation by PHE1. Helitrons containing PHE1 binding motifs are present in the promoter regions of several PHE1-targeted orthologs in the Brassicaceae, indicating ancestral insertion events [35]. It is tempting to speculate that the distribution of type I MADS-box transcription factor binding motifs by helitron TEs contributed to endosperm evolution by allowing the recruitment of crucial developmental genes into a common transcriptional network, analogous to the TE-driven diversification of the mammalian placenta (Figure 1).

Besides the helitron-mediated amplification of binding motifs targeted by the type I MADS-box transcription factor PHE1, the binding motifs recognized by the type II MADS-box transcription factor SEPALLATA3 (SEP3) were transposed by TEs (Figure 2), which was initially recognized in *A. lyrata* [36] and, based on a new method to detect degenerated TE sequences, also found in *A. thaliana* [37]. This method revealed large numbers of remnant TEs in *A. thaliana* which are close to genes and enriched for transcription factor binding motifs, in agreement with the hypothesis that domesticated regions of TEs are less likely to decay [37]. Importantly, some of the TE-derived regulatory elements are conserved across rosids and overrepresented in the flowering gene regulatory network, suggesting a deeply conserved functional role for these sequences since the spread of flowering plants in the Cretaceous [37]. In contrast to *A. thaliana*, in *A. lyrata*, many TEs with SEP3 recognized motifs are not yet decayed and could therefore be detected [36], suggesting recent, lineage-specific transposition events. Indeed, a repetitive element family similar to Copia-type LTR retrotransposons containing SEP3 bound motifs are present with only nine copies in *A. thaliana*, contrasting with more than 169 copies in *A. lyrata*. It is nevertheless also possible that this repeat family was lost in *A. thaliana* [36]. Genes being specifically controlled by TE-transposed SEP3 binding motifs in *A. lyrata* have potential functional roles in reproductive development [36]; whether they determine morphological differences between both species like flower organ size, remains an attractive hypothesis to be tested.

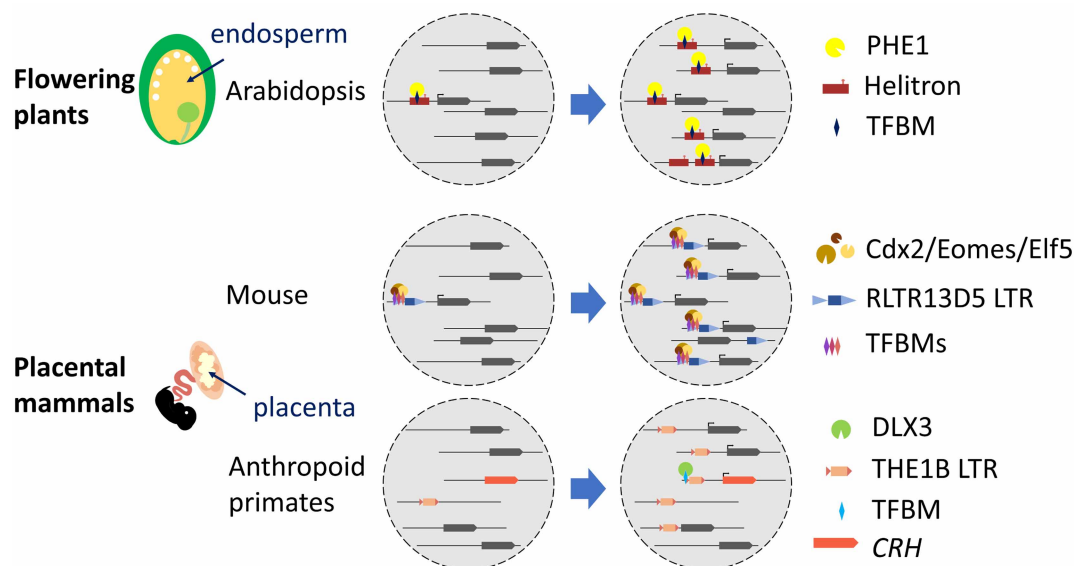


Figure 1. Convergent evolution of the mammalian placenta and the *Arabidopsis* endosperm.

In the *Arabidopsis* lineage, an ancestral helitron containing transcription factor binding motifs (TFBMs) for the PHE1 MADS-box transcription factor transposed into the promoter regions of other genes, transferring the TFBM and bringing those genes under the regulatory control of PHE1 [35]. In mammals, the mouse ancestral genome, after divergence from the rat lineage, acquired an LTR retrotransposon harboring three TFBMs in the LTR region recognized by Cdx2/Eomes/Elf5. This LTR family, *RLTR13D5*, became activated during mouse lineage evolution and subsequently distributed the three TFBMs into multiple locations, recruiting lineage-specific targets into the placental regulatory network [33]. In the anthropoid primate lineage, a different LTR family, *THE1B*, became amplified and plays a regulatory role for a suite of downstream genes. Noticeably, the LTR edge of one *THE1B* member joined with local genomic sequences and gave rise to a DLX3 binding motif, which is recognized by DLX3 and activates the key placental gene *CRH* in primates [34].

Transposition of MADS-box transcription factor binding motifs is not exceptional. A recent study proposed that TE transposition underpins the evolution of C_4 photosynthesis among grasses [38]. C_4 photosynthesis repeatedly evolved from C_3 photosynthesis and is characterized by higher efficiency in light, water, and nitrogen usage [39–41]. This is achieved by a special leaf morphology, two types of chloroplasts, as well as a series of enzymes specifically expressed in different leaf tissues of C_4 plants [40–42]. Based on a systematic genome-wide survey between maize and rice, it was found that the evolution of C_4 photosynthesis was associated with the recruitment of more than 1000 C_4 -associated motifs from pre-existing *cis*-regulatory elements originally belonging to non-photosynthetic genes into photosynthetic genes. Strikingly, more than 60% of the potentially recruited motifs are located in TEs, suggesting TE transposition has facilitated the distribution of C_4 -associated motifs. Both retrotransposons and DNA transposons, in particular short miniature inverted-repeat TEs (MITEs) were involved in this process (Figure 3). Strikingly, despite the large differences between C_3 and C_4 photosynthesis, C_4 photosynthesis has independently and rapidly evolved in more than 60 lineages [40]. The dispersal of pre-existing motifs by TEs provides a compelling mechanism underpinning the repeated evolution of C_4 photosynthesis. Importantly, C_4 photosynthesis is an adaptation of the C_3 pathway to dry environments by allowing to improve photosynthetic efficiency and minimizing water loss [40,41,43]. Therefore, the TE-driven evolution of C_4 photosynthesis provides a striking example how TE transposition benefits host adaptation.

TEs promote adaptation to environmental stresses

The response to stress is shaped by natural selection, leading to changes in the transcriptional regulation of a variety of genes. Based on her fundamental work in maize, Barbara McClintock was the first one to propose that stress activates TEs, which generates mutations that promote the adaptation to stress [44]. It is now generally accepted that TEs indeed play a key role in translating environmental cues into changes at the genomic level (Figure 4). The ability of TEs to respond to stress is in many cases mediated by *cis*-regulatory elements

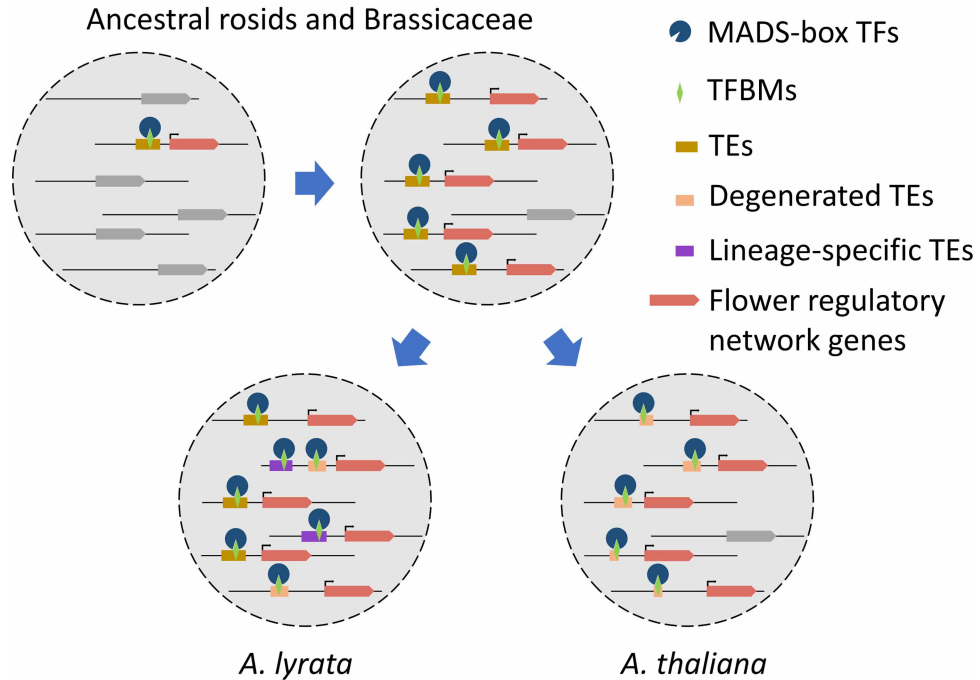


Figure 2. TEs rewired the regulatory network modulating flower development through transposition during rosid evolution.

Type II MADS-box transcription factors (TFs), including SEP3, are a major group of TFs regulating flower development. Their binding motifs (TFBMs) have been amplified by several TE families during the evolution of rosids and Brassicaceae lineages. New genes are continuously recruited into the flower regulatory network by novel TFBMs located in identified TE sequences. Some of the ancestral TEs have decayed, while the TFBMs were retained when functionally relevant [37]. Some TEs recently amplified in *A. lyrata* and generated novel targets, but not *A. thaliana* [36].

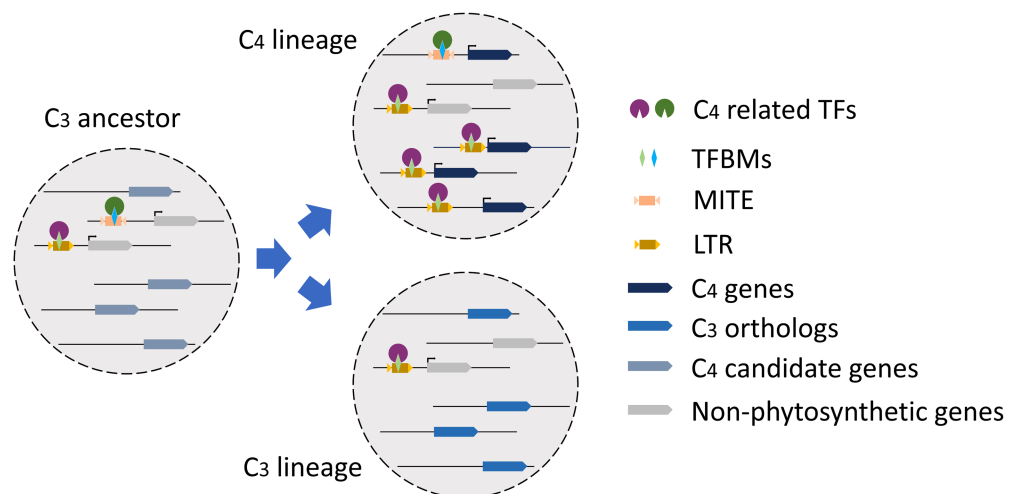


Figure 3. TEs promote the transition from C₃ to C₄ photosynthesis by transposition of transcription factor binding motifs.

TEs distributed C₄-related transcription factor binding motifs (TFBMs) from pre-existing *cis*-regulatory elements of non-photosynthetic genes, generating a novel regulatory network of C₄ photosynthetic genes [38].

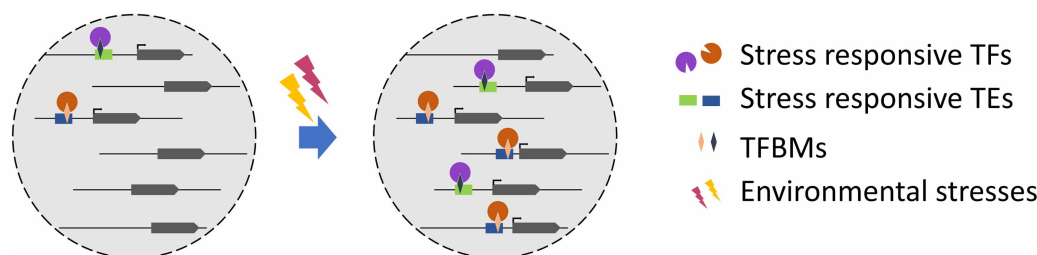


Figure 4. TEs contribute to stress adaptation.

TEs with stress-responsive transcription factor binding motifs (TFBMs) can amplify themselves in response to environmental stresses, thereby distributing TE-born responsive elements to new locations and generating new targets controlled by the stress responsive transcription factors (TFs).

present in the LTR regions of retrotransposons that trigger expression of TEs in response to a particular stimulus [45]. Several retrotransposons contain regulatory sequences with similarity to well-characterized motifs required for the activation of stress-responsive genes [46–51]. Importantly, the insertion of LTR TEs nearby genes can also confer stress-mediated activation to such genes [52,53], revealing that gene promoters co-opted sequences from LTRs for their own regulation.

In addition to LTR retrotransposons, DNA transposons such as MITEs also have been demonstrated to confer stress-responsive transcription to proximal genes, such as the highly proliferative *mPing* MITE in rice, which controls a stress-inducible network of genes conferring tolerance to cold and salt [54]. *mPing* is likely to function as an enhancer element, since it is able to act regardless of whether it is inserted at the 5' or the 3' region of target genes. MITEs have rapidly and massively amplified throughout eukaryotic genomes, suggesting that similar stress-responsive networks have evolved in other organisms as well [54].

In maize, different TE families are associated with stress-responsive gene expression, including all major TE families, such as TIR DNA transposons, Gypsy and Copia LTR elements, and LINEs [55]. Most of the TEs associated with stress-responsive genes contain the consensus motif for the stress-responsive DREB/CBF transcription factors. Notably, allelic variation for TE insertions is strongly associated with variation in stress-responsive gene expression, linking TEs to adaptive stress responses [55].

TEs do not only confer stress-responsive gene activation, but can potentially amplify the stress response by starting to proliferate under stress. For example, the tomato Copia LTR retrotransposon family *Rider* accumulates transcripts and transposition intermediates in the form of extrachromosomal DNA in response to drought stress [56]. There are several *cis*-regulatory elements in the LTR regions of *Rider* linked to abscisic acid (ABA) signaling that may confer drought responsive expression to neighboring genes. It is tempting to speculate that drought-responsive transcriptional activation induces transposition and by that amplifies drought-responsive regulatory elements throughout the genome. Similarly, *ONSEN*, another Copia retrotransposon in *A. thaliana*, has acquired a heat-responsive element that is recognized by heat shock transcription factors (HSFs), resulting in heat-responsive transcription and production of extrachromosomal DNA that can successfully integrate in epigenetically compromised mutants [52,57]. Heat-responsiveness in the Brassicaceae *COPIA* families has recurrently evolved, suggesting an adaptive advantage [53].

That TEs contain binding motifs for stress-response transcription factors and regulate a wide spectrum of stress-responsive genes is a conserved phenomenon and also found in yeast and animals. In the fission yeast *Schizosaccharomyces pombe*, the LTR retrotransposon *Tf1* frequently targets stress response genes and enhances the expression of downstream genes, likely through the recruitment of ATF/CREB stress response factors that are modulators of the heat shock factor (HSF) transcription complex [58,59]. In the fruitfly *Drosophila melanogaster*, binding motifs for several transcription factors related to a variety of stress responses have been amplified by TEs; among them is also HSF [60]. Also, in *Caenorhabditis* nematodes, the binding motifs for HSF were transposed by species-specific helitrons [61]. In human and likely other primate genomes, *Alu* repeats, which occupy large proportions of primate genomes, harbor HSF binding motifs and regulate the heat shock response through an antisense-mediated mechanism [62]. Considering the conserved and critical role of HSFs in the regulation of the heat-response, the co-option of TE-derived heat shock elements seems deeply conserved since the origin of eukaryotes. Gene regulatory networks underlying the response to other stresses can

potentially have the same evolutionary history. The transcriptional network underlying innate immunity has also been rewired through TE domestication. *MER41-like* LTR retrotransposons have dispersed the binding motifs of several interferon-inducible transcription factors and thus reshaped the immunity regulatory network independently during their colonization of multiple mammalian genomes [63]. These examples demonstrate that the transcriptional networks responsive to environmental stresses, both abiotic and biotic, have been driven by TEs throughout eukaryote evolution.

TEs co-ordinate biosynthetic genes

Biosynthetic gene duplication and functional specialization is an important strategy of plant defense responses. However, how duplicated genes are wired into existing regulatory networks remained a long-standing open question. TEs can coordinate biosynthetic genes in the same metabolic pathway synchronized by a single master regulator. The TE-mediated recruitment of the duplicated gene *CYP82C2* into the indole-3-carbonitrile (ICN) biosynthetic pathway provides a nice example addressing this question (Figure 5). The ICN pathway produces tryptophan-derived defence metabolites upon pathogen attack in *A. thaliana*. The LINE retrotransposon (*EPCOT3*) transferred a WRKY33-binding motif close to *CYP82C2*, the key enzyme for the last step in 4-hydroxy-ICN biosynthesis. WRKY33 directly activates the expression of other genes in this pathway, thus, the control of *CYP82C2* by WRKY33 was the required key step for the production of the inducible defense metabolite 4-hydroxy-ICN [64]. We speculate that there are probably many other cases where the TE-mediated recruitment of a newly duplicated gene into a regulatory network enriched chemical diversity, producing lineage-specific metabolites adaptive for plant immunity.

Regulatory motifs potentially facilitating DNA TE transpositions

The E2F family of transcription factors has a broad spectrum of functions in the regulation of cell cycle related processes, including DNA repair, cell proliferation, and differentiation in both animals and plants [65,66]. In line with the fundamental functions controlled by E2F transcription factors, E2Fs are highly conserved among organisms, as are the targeted binding motifs [67]. Despite the fundamental function of E2Fs and high level of conservation of this pathway, E2F motifs were extensively amplified in the Brassicaceae [68]. In *A. thaliana*, more than 70% of E2F binding motifs locate in TEs, especially MITEs. Interestingly, out of all potential E2F recognition motifs fitting the E2F consensus sequence, only one of them is overrepresented in TEs and dominant in the genome, evidencing it as the motif amplified through TE transposition. The capture of the E2F motifs by TEs likely occurred in an ancestral Brassicaceae genome and was then amplified to different extend

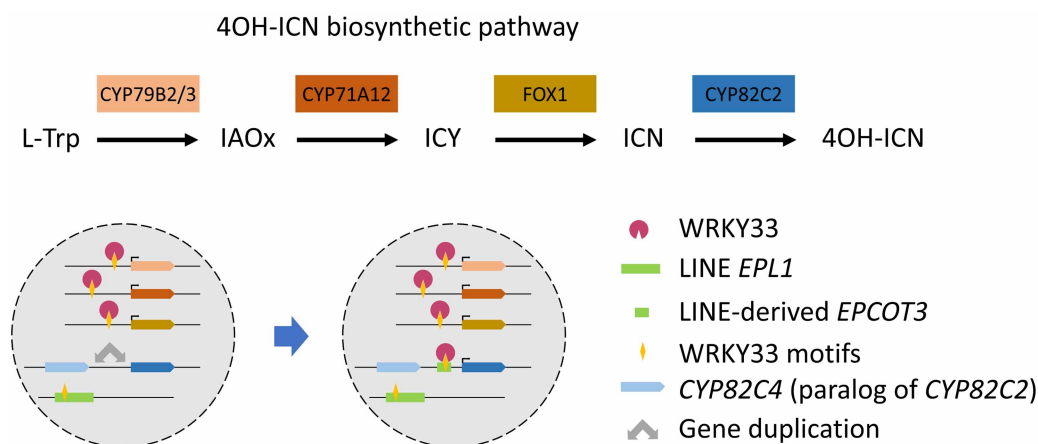


Figure 5. A LINE retrotransposon allowed WRKY33 to regulate *CYP82C2* in response to pathogen infection.

CYP82C2 and *CYP82C4* are tandem duplicates with conserved biochemical function. After gene duplication, *CYP82C2* in *Arabidopsis thaliana* acquired WRKY33-related motifs by insertion of a LINE retrotransposon. This generated a new pathogen defense pathway controlled by WRKY33 [64,106].

during Brassicaceae evolution. Nevertheless, while many of the transposed E2F motifs are bound by E2F, only few of them are close to promoters and therefore of potential regulatory relevance [68]. This conclusion is supported by a study identifying binding motifs for the E2F-interacting cell-fate master regulator, RETINOBLASTOMA RELATED 1 (RBR1) [69]. RBR1 binds to E2F motifs that have been amplified by TEs, but apparently does not regulate genes downstream of those E2F-TE motifs [69]. Therefore, whether binding of E2F/RBR1 to distally located transposed motifs has regulatory relevance seems unlikely. Instead of being of functional relevance, it is possible that facilitated transposition of E2F motifs is a consequence of E2F/RBR1 binding to those sites, which may promote DNA replication and/or repair and thus facilitate transposition of DNA TEs [69].

TEs facilitate transcription factor network evolution

The emergence of transcription factor binding motifs by *de novo* accumulation of mutations can occur on a small evolutionary timescale; time estimates for a given 5-mer to emerge in at least one of all human promoters are around 7500 years, for 8-mers around 350 000 years and for 10-mers around 700 000 to 4.8 million years, depending on the sequence context [70]. However, the actual binding of transcription factors may not simply rely on the presence of the motifs, but on additional specificity determinants outside the core motifs. This can include additional sequence features, other transcription factor binding motifs, or the general chromatin landscape determining access of transcription factors to their binding motifs [71–73]. Furthermore, once a transcription factor binding motif and the appropriate surrounding environment have evolved, in order for this transcription factor to control a network of genes, these binding motifs need to spread through the genome. This process has likely been facilitated by TEs. If TEs insert close to pre-existing transcription factor binding motifs, they can transpose these motifs to new locations, generating new transcriptional networks controlled by these transcription factors [6,38,74–76]. This process can explain the rapid and repeated evolution of transcriptional networks, as discussed in the previous sections.

The regulatory impact of TEs transposing binding motifs depends on their position of insertion. Many TEs have evolved mechanisms to favor integration sites that maximize their chance of propagation. This can be gene-poor heterochromatic regions for some TE families, or open, transcriptionally active regions for others [23,77]. Examples of TEs inserting favorably in transcriptionally active regions in plants are *Mutator* elements in maize [78] and *mPing* MITEs in rice [54], while helitrons preferentially insert in intergenic regions [79] and LTR retrotransposons show a higher preference to insert into older copies of the same family [80]. Nevertheless, it is important to note that selection may obscure initial insertion preferences, so the picture we have now may be biased [23].

It is predictable that this process occurs more frequently in lineages with higher transposition activities. In the case of SEP3, a much higher proportion of SEP3-bound sequence motifs was located in transposons in *A. lyrata* than in *A. thaliana* [36]. This coincides with the observation that TE activity is much higher in *A. lyrata* than in *A. thaliana* [81]. Thus, the advantages of effective TE domestication can potentially balance out the negative impact of TE activities.

The detection of TE-mediated amplification of transcription factor binding motifs

Current knowledge of TE-associated transcription factor binding motifs is mainly derived from two approaches. Some studies first experimentally identify transcription factor binding motifs (e.g. by chromatin immunoprecipitation followed by sequencing) and then test for the enrichment of these motifs in TEs, while other studies directly scan TEs for the sequences of known transcription factor binding motifs [35,36,82–84]. The first approach provides the actual picture of transcription factor binding and the realized regulatory network. The second approach requires subsequent validation to rule out false positives, as many putative motifs are likely not utilized [35,36,68]. Nevertheless, the binding of a transcription factor to motifs does not necessarily imply a regulatory role. Functional validation is required to prove the relevance of specific motifs. Genome-editing tools such as the CRISPR/Cas9 system allow to test the functional relevance of TEs and binding motifs on putative target genes as demonstrated by [85,86].

The contribution of TEs to the genomic landscape of known transcription factor binding motifs has been well-studied in human [82,83,87]. A comparative genomic analysis of orthologous transcription factors and their binding motifs between human and mouse genomes revealed substantial TE-mediated regulatory network

divergence [88]. A pioneer study in plants comparing MITE-associated transcription factors in the two closely related species peach and *Prunus mume* came to a similar conclusion; in both species MITEs were transposing the same transcription factor binding motifs, but the extend of this translocation and the generated networks differed [84].

Importantly however, many TE-transposed transcription factor motifs have probably escaped detection, because the TE sequences degenerate over time making their identification difficult [37]. Therefore, the majority of TE-derived binding motifs that can be easily detected are likely to be young, and many of them may be transient over evolutionary time. Most of them will be functionally neutral and disappear during species evolution, and only those conveying fitness advantage will remain. Thus, conserved TE-derived binding motifs are likely to reveal key elements of transcriptional networks.

A comprehensive knowledge of ancestral TEs is required to fully appreciate the extent to which TEs contributed to generate transcriptional networks and thus contributed to developmental novelties. Two recent studies identified ancient TEs by using novel computational approaches detecting degenerated TE sequences. This elevated TE-derived sequences in the *A. thaliana* genome from the currently annotated ~20% to as much as 50% [37,89]. Compared with previously identified, putatively young TEs, remnant sequences of ancient TEs are shorter, but enriched for transcription factor binding motifs as consequence of functional conservation [37]. Some of the TE-derived binding motifs are deeply conserved across rosids, indicating the shared fundamental regulatory pathways established by ancestral TE activities. Some remnant TEs contain multiple binding motifs associated with different transcription factors, evidencing the complexity of regulatory networks built by TEs [37]. These observations strongly support the role of TEs in rewiring gene regulatory networks with long-term evolutionary impact.

The evolutionary significance of TE-mediated dispersal of transcription factor binding motifs

As outlined above, TEs are, undeniably, major contributors to developmental innovation and adaptive responses to environmental conditions. What has been underappreciated thus far, is the potential role of TE-mediated transfer of transcription factor binding motifs in concert with gene duplication. After gene duplication, there are many potential trajectories for regulatory divergence. Regulatory neofunctionalization contributes to the retention of duplicated genes, where one duplicate can acquire novel regulatory elements and subsequently adopt new expression profiles [90–92]. Allopolyploidization, the merger of two genomes due to hybridization followed by genome doubling gives rise to large numbers of gene duplicates and is accompanied by relaxed TE repression and bursts of novel TE insertions [16,93]. Autopolyploidization leads to whole genome duplication without hybridization and transcriptome shock, nevertheless, it generates many gene duplicates with relaxed purifying selection allowing TE over-accumulation [94]. Thus, there seems to be a large potential for TEs in distributing transcription factor binding motifs and driving divergence and functional retention of homeologs in polyploids that remains to be investigated. Besides polyploidization, many small-scale duplication events, such as tandem duplications, produce targets for novel TE insertions that convey new regulatory functions. As discussed above, the WRKY33-binding site in the TE-derived *EPCOT3* contributed to the regulatory neofunctionalization of the duplicated gene *CYP82C2* in *A. thaliana*, rendering it non-redundant with other *CYP82C* paralogs in the tandem cluster [64]. Similarly, in the regulatory network of PHE1 in the *A. thaliana* endosperm, some of the helitron-recruited PHE1 binding motifs are in recent duplicates in the Brassicaceae lineage that diverged from their paralogs [35]. The contribution of TEs to the retention and divergence of duplicated genes could be an interesting topic of further study.

Many signaling components and transcription factors such as HOX and TCPs that control organ patterning in animals and plants, respectively, have been repeatedly recruited to function in the regulation of comparable morphological traits or similar developmental patterns. This type of evolutionary convergence has been termed ‘deep homology’, and the genes are referred to as ‘evolutionary toolkits’ [95–97]. An intriguing example is the transcription factor family CYCLOIDEA2 in flowering plants that has been repeatedly recruited for the establishment of bilateral flower symmetry from the ancestral radial state [98,99]. Based on the discussed ability of TEs to transpose transcription factor binding motifs, we propose that TEs play the role of evolutionary toolkits by transferring binding motifs of major developmental regulators. This could explain repeated and fast evolving convergent phenotypic innovation like bilateral flower asymmetry or C_4 photosynthesis in a lineage-specific

manner. Thus, TEs are powerful non-coding toolkits that construct complex developmental regulatory networks and thus contribute to species adaptation and radiation throughout the evolutionary history of all eukaryotes.

If TEs act as evolutionary toolkits, is it possible to take advantage of them as novel genetic tools for crop improvement? For example, the application of inducible *mPing* could potentially accelerate stress-tolerance breeding in rice [100]. Likewise, a synthetic stress-activated TE with a binding motif for a stress-responsive transcription factor could possibly generate novel candidate targets regulated in response to stress. By chemical suppression of the epigenetic silencing machinery, rice LTR retrotransposons such as *Houba* and *Tos17* could be successfully mobilized, leading to novel insertions [101–103]. TEs are known to have facilitated crop domestication naturally. A famous example is the retrotransposon *Hopscotch*, which inserted in the upstream region of the well-known maize domestication gene *teosinte branched1 (tb1)* and acts as an enhancer of gene expression. This promoted increased apical dominance in maize compared with its progenitor, generating a preferred phenotype for agricultural purpose [104]. We propose that by controlled unleashing of specific TEs, novel and improved traits could be generated in a much shorter time than conventional breeding could achieve. Thus, domestication by domesticated TEs offers large future potential.

Perspectives

- Importance of the field: Previously considered as genomic junk, TEs become increasingly recognized as major drivers of genome evolution. Their wide presence in prokaryotes and eukaryotes and maintenance over long evolutionary timescales indicates that TEs have important regulatory roles, which have started to be unraveled over recent years.
- Current thinking: TEs frequently contain regulatory elements that are recognized by transcription factors. Mobile TEs can transfer these regulatory sites, amplifying them through the genome and thus recruiting nearby genes into the regulatory network of the corresponding transcription factors. This process occurred repeatedly throughout the evolutionary history of diverse lineages, giving rise to a wide spectrum of phenotypic novelties, promoting adaptation and diversification.
- Future directions: That TEs generate transcriptional networks and promote developmental novelties is supported by several key studies. However, to fully appreciate the extent to which TEs promoted evolutionary novelties, a systematic investigation of TE-transferred transcription factor binding motifs and their functional relevance remains an important future endeavor. Newly developed bioinformatics tools to detect TE remnants [37] and improved tools to map transcription factor binding motifs [105] provide the technical base to approach these exciting questions.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by a grant from the Swedish Research Council (to C.K.) and support from the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (to C.K.).

Abbreviations

ABA, abscisic acid; HSF, heat shock factor; ICN, indole-3-carbonitrile; LINEs, long interspersed nuclear elements; LTRs, long terminal repeats; MITEs, miniature inverted-repeat TEs; RBR1, RETINOBLASTOMA RELATED 1; SEP3, SEPALLATA3; SINEs, short interspersed nuclear elements; TEs, transposable elements; TIRs, terminal inverted repeats.

References

- 1 Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G. et al. (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762 <https://doi.org/10.1101/gr.080663.108>
- 2 Bennetzen, J.L. and Wang, H. (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530 <https://doi.org/10.1146/annurev-arplant-050213-035811>
- 3 El Baidouri, M. and Panaud, O. (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol. Evol.* **5**, 954–965 <https://doi.org/10.1093/gbe/evt025>
- 4 Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368 <https://doi.org/10.1146/annurev.genet.40.110405.090448>
- 5 Arkhipova, I.R. (2017) Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mob. DNA* **8**, 19 <https://doi.org/10.1186/s13100-017-0103-2>
- 6 Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C. et al. (2016) Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat. Commun.* **7**, 10716 <https://doi.org/10.1038/ncomms10716>
- 7 Grabundzija, I., Hickman, A.B. and Dyda, F. (2018) Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. *Nat. Commun.* **9**, 1278 <https://doi.org/10.1038/s41467-018-03688-w>
- 8 Xiong, W., Dooner, H.K. and Du, C. (2016) Rolling-circle amplification of centromeric Helitrons in plant genomes. *Plant J.* **88**, 1038–1045 <https://doi.org/10.1111/tpj.13314>
- 9 Lisch, D. (2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49–61 <https://doi.org/10.1038/nrg3374>
- 10 Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B. et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* **8**, 973–982 <https://doi.org/10.1038/nrg2165>
- 11 Deniz, Ö., Frost, J.M. and Branco, M.R. (2019) Regulation of transposable elements by DNA modifications. *Nat Rev Genet.* **20**, 417–431 <https://doi.org/10.1038/s41576-019-0117-3>
- 12 Gutbrod, M.J. and Martienssen, R.A. (2020) Conserved chromosomal functions of RNA interference. *Nat. Rev. Genet.* **21**, 311–331 <https://doi.org/10.1038/s41576-019-0203-6>
- 13 Wang, Z. and Baulcombe, D.C. (2020) Transposon age and non-CG methylation. *Nat. Commun.* **11**, 1221 <https://doi.org/10.1038/s41467-020-14995-6>
- 14 Gouil, Q. and Baulcombe, D.C. (2016) DNA methylation signatures of the plant chromomethyltransferases. *PLoS Genet.* **12**, e1006526 <https://doi.org/10.1371/journal.pgen.1006526>
- 15 Yu, C.W., Tai, R., Wang, S.C., Yang, P., Luo, M., Yang, S. et al. (2017) HISTONE DEACETYLASE6 acts in concert with histone methyltransferases SUVH4, SUVH5, and SUVH6 to regulate transposon silencing. *Plant Cell* **29**, 1970–1983 <https://doi.org/10.1105/tpc.16.00570>
- 16 Oliver, K.R., McComb, J.A. and Greene, W.K. (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol Evol.* **5**, 1886–1901 <https://doi.org/10.1093/gbe/evt141>
- 17 de Souza, F.S., Franchini, L.F. and Rubinstein, M. (2013) Exaptation of transposable elements into novel *cis*-regulatory elements: is the evidence always strong? *Mol. Biol. Evol.* **30**, 1239–1251 <https://doi.org/10.1093/molbev/mst045>
- 18 Cui, X. and Cao, X. (2014) Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Curr. Opin. Plant Biol.* **21**, 83–88 <https://doi.org/10.1016/j.pbi.2014.07.001>
- 19 Hirsch, C.D. and Springer, N.M. (2017) Transposable element influences on gene expression in plants. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1860**, 157–165 <https://doi.org/10.1016/j.bbagr.2016.05.010>
- 20 Mita, P. and Boeke, J.D. (2016) How retrotransposons shape genome regulation. *Curr. Opin. Genet. Dev.* **37**, 90–100 <https://doi.org/10.1016/j.gde.2016.01.001>
- 21 Cowley, M. and Oakey, R.J. (2013) Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet.* **9**, e1003234 <https://doi.org/10.1371/journal.pgen.1003234>
- 22 Elbarbary, R.A., Lucas, B.A. and Maquat, L.E. (2016) Retrotransposons as regulators of gene expression. *Science* **351**, aac7247 <https://doi.org/10.1126/science.aac7247>
- 23 Chuong, E.B., Elde, N.C. and Feschotte, C. (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 <https://doi.org/10.1038/nrg.2016.139>
- 24 Jangam, D., Feschotte, C. and Betrán, E. (2017) Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* **33**, 817–831 <https://doi.org/10.1016/j.tig.2017.07.011>
- 25 Britten, R.J. and Davidson, E.H. (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46**, 111–138 <https://doi.org/10.1086/406830>
- 26 Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405 <https://doi.org/10.1038/nrg2337>
- 27 Lynch, V.J., Leclerc, R.D., May, G. and Wagner, G.P. (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **43**, 1154–1159 <https://doi.org/10.1038/ng.917>
- 28 Kunarso, G., Chia, N., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S. et al. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631–634 <https://doi.org/10.1038/ng.600>
- 29 Notwell, J.H., Chung, T., Heavner, W. and Bejerano, G. (2015) A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat. Commun.* **6**, 6644 <https://doi.org/10.1038/ncomms7644>
- 30 Rebollo, R., Romanish, M.T. and Mager, D.L. (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **46**, 21–42 <https://doi.org/10.1146/annurev-genet-110711-155621>
- 31 Chuong, E.B. (2018) The placenta goes viral: Retroviruses control gene expression in pregnancy. *PLoS Biol.* **16**, e3000028 <https://doi.org/10.1371/journal.pbio.3000028>
- 32 Frank, J.A. and Schötte, C. (2017) Co-option of endogenous viral sequences for host cell function. *Curr. Opin. Virol.* **25**, 81–89 <https://doi.org/10.1016/j.coviro.2017.07.021>

- 33 Chuong, E.B., Rumi, M.A., Soares, M.J. and Baker, J.C. (2013) Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45**, 325–329 <https://doi.org/10.1038/ng.2553>
- 34 Dunn-Fletcher, C.E., Muglia, L.M., Pavlicev, M., Wolf, G., Sun, M.A., Hu, Y.C. et al. (2018) Anthropoid primate-specific retroviral element *THE1B* controls expression of CRH in placenta and alters gestation length. *PLoS Biol.* **16**, e2006337 <https://doi.org/10.1371/journal.pbio.2006337>
- 35 Batista, R.A., Moreno-Romero, J., Qiu, Y., van Boven, J., Santos-González, J., Figueiredo, D.D. et al. (2019) The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife* **8**, e50541 <https://doi.org/10.7554/eLife.50541>
- 36 Muiño, J.M., de Bruijn, S., Pajaro, A., Geuten, K., Vingron, M., Angenent, G.C. et al. (2016) Evolution of DNA-binding sites of a floral master regulatory transcription factor. *Mol. Biol. Evol.* **33**, 185–200 <https://doi.org/10.1093/molbev/msv210>
- 37 Baud, A., Wan, M., Nouaud, D., Francillonne, N., Anxolabéhère, D. and Quesneville, H. (2020) Traces of transposable elements in genome dark matter co-opted by flowering gene regulation networks. *bioRxiv* <https://doi.org/10.1101/547877>
- 38 Cao, C., Xu, J., Zheng, G. and Zhu, X.G. (2016) Evidence for the role of transposons in the recruitment of *cis*-regulatory motifs during the evolution of *C₄* photosynthesis. *BMC Genomics* **17**, 201 <https://doi.org/10.1186/s12864-016-2519-3>
- 39 Edwards, E.J. and Smith, S.A. (2010) Phylogenetic analyses reveal the shady history of *C₄* grasses. *Proc. Natl Acad. Sci. U.S.A.* **107**, 2532–2537 <https://doi.org/10.1073/pnas.0909672107>
- 40 Sage, R.F., Christin, P.-A. and Edwards, E.J. (2011) The *C₄* plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3155–3169 <https://doi.org/10.1093/jxb/err048>
- 41 Sage, R.F., Sage, T.L. and Kocacinar, F. (2012) Photorespiration and the evolution of *C₄* photosynthesis. *Annu. Rev. Plant Biol.* **63**, 19–47 <https://doi.org/10.1146/annurev-arplant-042811-105511>
- 42 Brown, W.V. (1975) Variations in anatomy, associations, and origins of Kranz tissue. *Am. J. Bot.* **62**, 395–402 <https://doi.org/10.1002/j.1537-2197.1975.tb14062.x>
- 43 Ehleringer, J.R., Sage, R.F., Flanagan, L.B. and Pearcy, R.W. (1991) Climate change and the evolution of C(4) photosynthesis. *Trends Ecol. Evol.* **6**, 95–99 [https://doi.org/10.1016/0169-5347\(91\)90183-X](https://doi.org/10.1016/0169-5347(91)90183-X)
- 44 McClintock, B. (1984) The significance of responses of the genome to challenge. *Science* **226**, 792–801 <https://doi.org/10.1126/science.15739260>
- 45 Kumar, A. and Bennetzen, J.L. (1999) Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532 <https://doi.org/10.1146/annurev.genet.33.1.479>
- 46 Beguiristain, T., Grandbastien, M.A., Puigdomènech, P. and Casacuberta, J.M. (2001) Three *Tnt1* subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiol.* **127**, 212–221 <https://doi.org/10.1104/pp.127.1.212>
- 47 Kimura, Y., Tosa, Y., Shimada, S., Sogo, R., Kusaba, M., Sunaga, T. et al. (2001) OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. *Plant Cell Physiol.* **42**, 1345–1354 <https://doi.org/10.1093/pccp/pce171>
- 48 Woodrow, P., Pontecorvo, G., Ciarmiello, L.F., Fuggi, A. and Carillo, P. (2011) Ttd1a promoter is involved in DNA-protein binding by salt and light stresses. *Mol. Biol. Rep.* **38**, 3787–3794 <https://doi.org/10.1007/s11033-010-0494-3>
- 49 Grandbastien, M.A., Audeon, C., Bonnivard, E., Casacuberta, J.M., Chalhoub, B., Costa, A.P. et al. (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110**, 229–241 <https://doi.org/10.1159/000084957>
- 50 Salazar, M., González, E., Casaretto, J.A., Casacuberta, J.M. and Ruiz-Lara, S. (2007) The promoter of the TLC1.1 retrotransposon from *solanum chilense* is activated by multiple stress-related signaling molecules. *Plant Cell Rep.* **26**, 1861–1868 <https://doi.org/10.1007/s00299-007-0375-y>
- 51 Galindo-González, L., Mhiri, C., Deyholos, M.K. and Grandbastien, M.A. (2017) LTR-retrotransposons in plants: Engines of evolution. *Gene* **626**, 14–25 <https://doi.org/10.1016/j.gene.2017.04.051>
- 52 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 <https://doi.org/10.1038/nature09861>
- 53 Pietzenek, B., Markus, C., Gaubert, H., Bagwan, N., Merotto, A., Bucher, E. et al. (2016) Recurrent evolution of heat-responsiveness in brassicaceae COPIA elements. *Genome Biol.* **17**, 209 <https://doi.org/10.1186/s13059-016-1072-3>
- 54 Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O. et al. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **461**, 1130–1134 <https://doi.org/10.1038/nature08479>
- 55 Makarevitch, I., Waters, A.J., West, P.T., Stitzer, M., Hirsch, C.N., Ross-Ibarra, J. et al. (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **11**, e1004915 <https://doi.org/10.1371/journal.pgen.1004915>
- 56 Benoit, M., Drost, H.-G., Catoni, M., Gouil, Q., Lopez-Gomollon, S., Baulcombe, D. et al. (2019) Environmental and epigenetic regulation of *ridier* retrotransposons in tomato. *PLoS Genet.* **15**, e1008370 <https://doi.org/10.1371/journal.pgen.1008370>
- 57 Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M. and Mittelsten Scheid, O. (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet* **10**, e1004115 <https://doi.org/10.1371/journal.pgen.1004115>
- 58 Feng, G., Leem, Y.E. and Levin, H.L. (2013) Transposon integration enhances expression of stress response genes. *Nucleic Acids Res.* **41**, 775–789 <https://doi.org/10.1093/nar/gks1185>
- 59 Takii, R., Fujimoto, M., Tan, K., Takaki, E., Hayashida, N., Nakato, R. et al. (2015) ATF1 modulates the heat shock response by regulating the stress-inducible heat shock factor 1 transcription complex. *Mol. Cell. Biol.* **35**, 11–25 <https://doi.org/10.1128/MCB.00754-14>
- 60 Villanueva-Canas, J.L., Horvath, V., Aguilera, L. and Gonzalez, J. (2019) Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*. *Nucleic Acids Res.* **47**, 6842–6857 <https://doi.org/10.1093/nar/gkz490>
- 61 Garrigues, J.M., Tsu, B.V., Daugherty, M.D. and Pasquinelli, A.E. (2019) Diversification of the *Caenorhabditis* heat shock response by helitron transposable elements. *eLife* **8**, e51139 <https://doi.org/10.7554/eLife.51139>
- 62 Pandey, R., Mandal, A.K., Jha, V. and Mukerji, M. (2011) Heat shock factor binding in *Alu* repeats expands its involvement in stress through an antisense mechanism. *Genome Biol.* **12**, R117 <https://doi.org/10.1186/gb-2011-12-11-r117>
- 63 Chuong, E.B., Eldre, N.C. and Feschotte, C. (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087 <https://doi.org/10.1126/science.125497>
- 64 Barco, B., Kim, Y. and Clay, N.K. (2019) Expansion of a core regulon by transposable elements promotes *Arabidopsis* chemical diversity and pathogen defense. *Nat. Commun.* **10**, 3444 <https://doi.org/10.1038/s41467-019-11406-3>
- 65 van den Heuvel, S. and Dyson, N.J. (2008) Conserved functions of the pRB and E2F families. *Nat. Rev. Mol. Cell Biol.* **9**, 713–724 <https://doi.org/10.1038/nrm2469>

- 66 Lammens, T., Li, J., Leone, G. and De Veylder, L. (2009) Atypical E2Fs: new players in the E2F transcription factor family. *Trends Cell Biol.* **19**, 111–118 <https://doi.org/10.1016/j.tcb.2009.01.002>
- 67 DeGregori, J. and Johnson, D.G. (2006) Distinct and overlapping roles for E2F family members in transcription, proliferation and apoptosis. *Curr. Mol. Med.* **6**, 739–748 <https://doi.org/10.2174/1566524010606070739>
- 68 Hénaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C. et al. (2014) Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of *Brassica* species. *Plant J.* **77**, 852–862 <https://doi.org/10.1111/tpj.12434>
- 69 Bouyer, D., Heese, M., Chen, P., Harashima, H., Roudier, F., Grüttner, C. et al. (2018) Genome-wide identification of RETINOBLASTOMA RELATED 1 binding sites in arabidopsis reveals novel DNA damage regulators. *PLoS Genet.* **14**, e1007797 <https://doi.org/10.1371/journal.pgen.1007797>
- 70 Behrens, S. and Vingron, M. (2010) Studying the evolution of promoter sequences: a waiting time problem. *J. Comput. Biol.* **17**, 1591–1606 <https://doi.org/10.1089/cmb.2010.0084>
- 71 Guertin, M.J. and Lis, J.T. (2013) Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr. Opin. Genet. Dev.* **23**, 116–123 <https://doi.org/10.1016/j.gde.2012.11.008>
- 72 Levo, M., Zalckvar, E., Sharon, E., Dantas Machado, A.C., Kalma, Y., Lotam-Pompan, M. et al. (2015) Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* **25**, 1018–1029 <https://doi.org/10.1101/gr.185033.114>
- 73 Grossman, S.R., Zhang, X., Wang, L., Engreitz, J., Melnikov, A., Rogov, P. et al. (2017) Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc. Natl Acad. Sci. U.S.A.* **114**, E1291–E1300 <https://doi.org/10.1073/pnas.1621150114>
- 74 Pritham, E.J. and Feschotte, C. (2007) Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc. Natl Acad. Sci. U.S.A.* **104**, 1895–1900 <https://doi.org/10.1073/pnas.0609601104>
- 75 Thomas, J., Phillips, C.D., Baker, R.J. and Pritham, E.J. (2014) Rolling-circle transposons catalyze genomic innovation in a Mammalian lineage. *Genome Biol. Evol.* **6**, 2595–2610 <https://doi.org/10.1093/gbe/evu204>
- 76 Jiang, N., Bao, Z., Zhang, X., Eddy, S.R. and Wessler, S.R. (2004) Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **431**, 569–573 <https://doi.org/10.1038/nature02953>
- 77 Levin, H.L. and Moran, J.V. (2011) Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12**, 615–627 <https://doi.org/10.1038/nrg3030>
- 78 Liu, S., Yeh, C.T., Ji, T., Ying, K., Wu, H., Tang, H.M. et al. (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet.* **5**, e1000733 <https://doi.org/10.1371/journal.pgen.1000733>
- 79 Hu, K., Xu, K., Wen, J., Yi, B., Shen, J., Ma, C. et al. (2019) Helitron distribution in Brassicaceae and whole Genome Helitron density as a character for distinguishing plant species. *BMC Bioinformatics* **20**, 354 <https://doi.org/10.1186/s12859-019-2945-8>
- 80 Jedlicka, P., Lexa, M., Vanat, I., Hobza, R. and Kejnovsky, E. (2019) Nested plant LTR retrotransposons target specific regions of other elements, while all LTR retrotransposons often target palindromes and nucleosome-occupied regions: in silico study. *Mob. DNA* **10**, 50 <https://doi.org/10.1186/s13100-019-0186-z>
- 81 Hollister, J.D., Smith, L.M., Guo, Y.L., Ott, F., Weigel, D. and Gaut, B.S. (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl Acad. Sci. U.S.A.* **108**, 2322–2327 <https://doi.org/10.1073/pnas.1018222108>
- 82 Jacques, P.-É., Jeyakani, J. and Bourque, G. (2013) The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* **9**, e1003504 <https://doi.org/10.1371/journal.pgen.1003504>
- 83 Polavarapu, N., Mariño-Ramírez, L., Landsman, D., McDonald, J.F. and Jordan, I.K. (2008) Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. *BMC Genomics* **9**, 226 <https://doi.org/10.1186/1471-2164-9-226>
- 84 Morata, J., Marín, F., Payet, J. and Casacuberta, J.M. (2018) Plant lineage-specific amplification of transcription factor binding motifs by Miniature Inverted-Repeat Transposable Elements (MITEs). *Genome Biol Evol.* **10**, 1210–1220 <https://doi.org/10.1093/gbe/evy073>
- 85 Fuentes, D.R., Swigut, T. and Wysocka, J. (2018) Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *eLife* **7**, e35989 <https://doi.org/10.7554/eLife.35989>
- 86 Todd, C.D., Deniz, Ö., Taylor, D. and Branco, M.R. (2019) Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *eLife* **8**, e44344 <https://doi.org/10.7554/eLife.44344>
- 87 Lynch, V.J., Nnamani, M.C., Kapusta, A., Brayer, K., Plaza, S.L., Mazur, E.C. et al. (2015) Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* **10**, 551–561 <https://doi.org/10.1016/j.celrep.2014.12.052>
- 88 Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P. et al. (2014) Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976 <https://doi.org/10.1101/gr.168872.113>
- 89 Maumus, F. and Quesneville, H. (2014) Ancestral repeats have shaped epigenome and genome composition for millions of years in *arabidopsis thaliana*. *Nat. Commun.* **5**, 4104 <https://doi.org/10.1038/ncomms5104>
- 90 Liu, S.L., Baute, G.J. and Adams, K.L. (2011) Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in *arabidopsis thaliana*. *Genome Biol. Evol.* **3**, 1419–1436 <https://doi.org/10.1093/gbe/evr114>
- 91 Hughes, T.E., Langdale, J.A. and Kelly, S. (2014) The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Res.* **24**, 1348–1355 <https://doi.org/10.1101/gr.172684.114>
- 92 Kassahn, K.S., Dang, V.T., Wilkins, S.J., Perkins, A.C. and Ragan, M.A. (2009) Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res.* **19**, 1404–1418 <https://doi.org/10.1101/gr.086827.108>
- 93 Vicent, C.M. and Casacuberta, J.M. (2017) Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* **120**, 195–207 <https://doi.org/10.1093/aob/mcx078>
- 94 Baduel, P., Quadrana, L., Hunter, B., Bombliès, K. and Colot, V. (2019) Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat. Commun.* **10**, 5818 <https://doi.org/10.1038/s41467-019-13730-0>
- 95 Gerhart, J. and Kirschner, M. (2007) The theory of facilitated variation. *Proc. Natl Acad. Sci. U.S.A.* **104**, 8582–8589 <https://doi.org/10.1073/pnas.0701035104>
- 96 Carroll, S.B. (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 <https://doi.org/10.1016/j.cell.2008.06.030>

- 97 Shubin, N., Tabin, C. and Carroll, S. (2009) Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 <https://doi.org/10.1038/nature07891>
- 98 Yang, X., Pang, H.B., Liu, B.L., Qiu, Z.J., Gao, Q., Wei, L. et al. (2012) Evolution of double positive autoregulatory feedback loops in CYCLOIDEA2 clade genes is associated with the origin of floral zygomorphy. *Plant Cell* **24**, 1834–1847 <https://doi.org/10.1105/tpc.112.099457>
- 99 Zhang, W., Kramer, E.M. and Davis, C.C. (2010) Floral symmetry genes and the origin and maintenance of zygomorphy in a plant-pollinator mutualism. *Proc. Natl Acad. Sci. U.S.A.* **107**, 6388–6393 <https://doi.org/10.1073/pnas.0910155107>
- 100 Yasuda, K., Ito, M., Sugita, T., Tsukiyama, T., Saito, H., Naito, K. et al. (2013) Utilization of transposable element *mPing* as a novel genetic tool for modification of the stress response in rice. *Mol. Breed* **32**, 505–516 <https://doi.org/10.1007/s11032-013-9885-1>
- 101 Thieme, M., Lanciano, S., Balzergue, S., Daccord, N., Mirouze, M. and Bucher, E. (2017) Inhibition of RNA polymerase II allows controlled mobilisation of retrotransposons for plant breeding. *Genome Biol.* **18**, 134 <https://doi.org/10.1186/s13059-017-1265-4>
- 102 Paszkowski, J. (2015) Controlled activation of retrotransposition for plant breeding. *Curr. Opin. Biotechnol.* **32**, 200–206 <https://doi.org/10.1016/j.copbio.2015.01.003>
- 103 Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K. et al. (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15**, 1771–1780 <https://doi.org/10.1105/tpc.012559>
- 104 Studer, A., Zhao, Q., Ross-Ibarra, J. and Doebley, J. (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**, 1160–1163 <https://doi.org/10.1038/ng.942>
- 105 Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G. et al. (2019) CUT&tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 <https://doi.org/10.1038/s41467-019-09982-5>
- 106 Barco, B. and Clay, N.K. (2020) Hierarchical and dynamic regulation of defense-responsive specialized metabolism by WRKY and MYB transcription factors. *Front Plant Sci.* **10**, 1775 <https://doi.org/10.3389/fpls.2019.01775>