The regulatory potential of transposable elements in plant defence pathways

Ivona Glavincheska (4919691)

MSc programme Environmental Biology

Department of Biology, Graduate School of Life Sciences,

Faculty of Science, Utrecht University

Utrecht, the Netherlands

Writing assignment (7.5 ECTS): Literature review

March ’22 – April ‘22

Supervision: Dr Michael Seidl

Universiteit Utrecht

ABSTRACT

Microbial pathogens are one of the primary threats to global food security. Despite this, most microbes do not successfully establish pathogenesis, owing to plants multi-layered defence pathways. Plant defence is dependent on a global transcriptional response initiated by pathogen recognition. Activation of transcriptional expression of defence genes is dependent on *cis*-regulatory elements, the *trans*-actors that interact with them and other factors such as chromatin accessibility and non-coding RNAs. Recent research has shown that aside from promoting structural variation, transposable elements (TEs) can also contribute to the regulation of endogenous gene expression. Transposons are mobile genetic elements that can replicate and change their position in the genome through a process called transposition. In this review, I highlight that TEs have been repeatedly co-opted for transcriptional regulation in plant defence pathways and discuss the mechanisms through which TEs participate in gene regulation. The mobility of TEs can provide *cis*-regulatory elements at locations of integration or redirect the cellular silencing machinery to influence the expression of protein-coding sequences. Furthermore, non-coding RNAs derived from TE sequences can function as *trans*-actors in guiding various transcriptional and post-transcriptional mechanisms to influence gene expression. The magnitude at which TEs can regulate gene expression and the diversity of these mechanisms underlines the importance of TEs in the evolution of gene regulatory networks. Manipulation of TE sequences could hence provide a powerful tool in creating regulatory variation and rewiring of plant defence pathways for optimal response to pathogens. Further research would determine whether this is a feasible strategy for applied solutions in agriculture.

LAYMANS SUMMARY

Like animals and humans, plants are constantly exposed to a range of disease-causing microbes. Microbial disease decreases plants' health, aesthetic and economic value resulting in billions of euros worth of agricultural losses. In conventional farming, huge amounts of chemicals are used to stop disease pandemics in agricultural fields, sometimes with limited success. So, finding ways to reliably grow plants that are not susceptible to disease is important not only for gardeners and farmers but also for consumers. One way to do this would be to optimise the plant's own defence system. Indeed, plants, like us and other animals also have an immune system that protects them from most disease-causing microbes. Plant defence processes are initiated when plants recognise the invading microbe and then activate a coordinated genetic response to stop or slow down the microbes' growth. To be able to boost plants' natural defence mechanisms we need to understand the initiation and later progression of the plant's response to disease-causing microbes. However, regulation of gene expression is complicated and recent technological advancements have only underlined how little we truly know about this topic. Aside from plant genes that make up the main genetic material, plants and other organisms contain jumping genes which function largely on their own. As the name would suggest, jumping genes can replicate and reinsert themselves at different locations in DNA sequences and this movement creates gene changes that most of the time are harmful. Recently, scientists have suggested that jumping genes might not always be bad and can sometimes support activation of genetic signalling. Here, I will look at the ways in which these jumping genes can influence gene activation during infection with disease-causing microbes. It is evident that jumping genes can benefit plants in some cases and this is dependent on the characteristics of the jumping gene itself but also plant-specific factors. Better understanding of these processes could contribute towards innovative techniques in producing more resistant plants.

Table of Contents

1. Introduction

Microbial pathogens, such as bacteria, fungi, and oomycetes are one of the primary threats to global food security (Savary *et al.*, 2019). Pathogen infection adversely affects plant fitness and consequently the yield of agriculturally relevant crops (Savary *et al.*, 2019). For instance, 20-30% of the annual yield from five major agricultural crops, such as wheat and rice, is lost to pathogens and pests (Savary *et al.*, 2019). Despite the magnitude of pathogen-associated crop losses, plants are highly effective at resisting infection and most plant-pathogen interactions do not result in successful pathogenesis. Morphological structures and constitutive production of antimicrobial compounds deter the entry and growth of most pathogenic microbes. Pathogens that successfully subvert this first line of defence can be recognised by cell surface pattern recognition receptors (PRRs) and intracellular nucleotide-binding, leucine rich receptors (NLRs), which then initiate a coordinated immune response intended to stop pathogenesis (Zhou and Zhang, 2020).

Pathogen recognition by host receptors induces dynamic reprogramming of endogenous gene expression that is essential in launching effective defence responses (Zhou and Zhang, 2020). These defence signalling networks can affect diverse aspects of plant health and physiology, and consequently their initiation can impact plant development and reproduction (Figueroa-Macías *et al.*, 2021). Therefore, constitutive activation of defence signalling is not advantageous and under unchallenged conditions plants repress defence pathways to maintain developmental functions (Figueroa-Macías *et al.*, 2021). An extensive understanding of the coordinated transcriptional activation and downstream signalling during pathogenesis can hence aid in breeding crops with optimal defence responses. Gene expression is primarily regulated by *cis*-regulatory elements and the transcription factors (TFs) and cofactors that interact with them (Zhou and Zhang, 2020; Schmitz, Grotewold and Stam, 2022). Further complexity to

transcriptional regulation is added by dynamic changes of chromatin accessibility and the involvement of endogenous non-coding RNAs (ncRNAs; Bhogireddy *et al.*, 2021; Hannan Parker, Wilkinson and Ton, 2022). Despite advances in the tools available to study these processes, the regulatory dynamics of gene transcription remains largely unexplored (Schmitz, Grotewold and Stam, 2022).

Transposons are mobile genomic elements with the capacity to replicate and change their position in the host genome through a process called transposition. Historically, transposable elements (TEs) have been considered to function at the expense of stability of the host genome as most transpositions can result in detrimental mutations. However, an increasing body of research now recognizes TEs as major drivers of genome size evolution and genomic variation (Negi, Rai and Suprasanna, 2016; Galindo-González *et al.*, 2017; Qiu and Köhler, 2020; Roquis *et al.*, 2021). Most recently, the regulatory role of TEs in expression of plant genes, particularly in stress responses has gained increased attention (Cho, 2018; Wyler *et al.*, 2020; Gill *et al.*, 2021). TEs have been implicated in regulating the outcome of plant-pathogen interactions and a central role for TEs has been implied in epigenetic control of short- and long-term plant immunity (Hannan Parker, Wilkinson and Ton, 2022). However, the magnitude at which TEs can affect defence gene expression has not been defined. This review (Appendix 1) aims to assess the scope of TE co-option in defence gene regulatory networks and determine the mechanisms of TE-dependent defence gene regulation. Particularly, throughout the review I seek to gain further understanding on what the structural and biochemical properties allow TEs to act as efficient gene transcriptional regulators. I highlight examples demonstrating that TEs can repeatedly acquire functional roles in plant defence signalling pathways and that these can impact pathogen induced gene expression dependent on sequence and chromatin factors.

2. Transposable elements in plant genomes

TEs are ubiquitous components of eukaryotic genomes and occupy a large proportion of an organism's genetic material. Dependent on whether they employ an RNA or DNA intermediate for their transposition, TEs can be classified as retrotransposons (Class I) or DNA transposons (Class II) (Fambrini *et al.*, 2020). The TE fraction in plant genomes can vary dependent on the plant species and genome size. (Fambrini *et al.*, 2020) Notably, a small number of highly expanded TE families dominate plant genomes, especially *Gypsy* and *Copia* long terminal repeat (LTR) retrotransposons (Fambrini *et al.*, 2020). For instance, in maize up to 85% of the genome is composed of TEs, 90% of which are classified in the *Gypsy* and *Copia* superfamilies (Zhang and Qi, 2019). Transposition is mediated by a suit of proteins that are typically directly encoded by TEs (Fambrini *et al.*, 2020). Nevertheless, TEs that lack some or all necessary enzymes for transpositions, so-called non-autonomous TEs, can still transpose in the presence of activated compatible autonomous elements (Fambrini *et al.*, 2020). Non-autonomous TEs are largely derived from autonomous elements which have lost the complete or partial transpositional machinery through accumulation of mutations or chromosomal rearrangements (Fambrini *et al.*, 2020).

Unconstrainted global activation of TEs could result in an overwhelmingly high rate of mutation and disrupt genetic stability by insertions and rearrangements that compromise coding sequences (Naito *et al.*, 2009; Erdmann and Picard, 2020). To diminish the mutagenic effect of TEs, the host genome has evolved multi-layered mechanisms to recognize and repress TE activity (Erdmann and Picard, 2020). TE silencing mechanisms rely heavily on chromatin modifications that result in the formation of heterochromatin. Particularly DNA methylation on cytosines and post-transcriptional modification of histones, such as methylation of histone 3 at lysine 9 (H3K9) are well correlated with gene silencing (Erdmann and Picard, 2020). Current understanding of the pathways involved in TE silencing in plants has largely been derived from research in the model plant *Arabidopsis thaliana* (hereafter Arabidopsis). Thus, the mechanistic explanation provided here is based on these concepts but is considered largely applicable in other angiosperms, with notable exceptions (Noshay, Crisp and Springer, 2018; Erdmann and Picard, 2020).

In plants, DNA methylation at TE loci can occur on cytosines in all sequence contexts. i.e., at symmetric CG, CHG, and asymmetric CHH (H is A, T, or C). Initiation of *de novo* methylation at transcriptionally active TEs is dependent on 21-22 nucleotide (nt) small interfering RNA (siRNA) biogenesis mediated by RDR6 and overlapping with post-transcriptional gene silencing (PTGS; Erdmann and Picard, 2020). Downstream of RDR6, *de novo* methylation in all cytosine contexts requires factors involved in the canonical RNA-directed DNA methylation pathway (RdDM) guided by the RDR6-derived siRNAs (Erdmann and Picard, 2020). Once established, methylation patterns can be conservatively maintained through cell division by DNA methyltransferases that function dependent on sequence and chromosomal context (Erdmann and Picard, 2020). In Arabidopsis, MET1 maintains CG methylation at all loci, whereas CMT2 can maintain CHG and CHH methylation patterns only in pericentromeric regions (Appendix 2; Erdmann and Picard, 2020). In the chromosomal arms CHG methylation is maintained by CMT3, whereas CHH context methylation re-establishment is dependent on DRM2 and canonical RdDM guided by 24-nt siRNAs (Erdmann and Picard, 2020). Canonical RdDM function is maintained by the activity of two RNA polymerases. Pol IV transcribes siRNA precursors that are processed into 24-nt siRNAs and subsequently preferentially loaded onto AGO4. The siRNA-AGO complex targets nascent scaffold transcripts from Pol V by sequence complementarity and recruits DRM2 DNA methyltransferase (Erdmann and Picard, 2020). In Arabidopsis, all DNA methyltransferases, apart from MET1, are directly or indirectly guided by histone modifications present at target loci. Post-transcriptional histone modifications can be additionally recruited at DNA methylated loci driven by histone methyltransferases such as SUVH4/5/6 (Erdmann and Picard, 2020). A plethora of other auxiliary proteins and transcriptional complexes can be further recruited to impact chromatin organization in the formation of heterochromatic structures. This multi-layered silencing mechanism allows flexible regulation of TEs that is dependent on multitude of self-regulating markers and the relative position of the TE within the genome.

Despite the presence of effective silencing pathways, transpositional activation of TEs can still be detected in various plant species dependent on endogenous and environmental factors. Large-scale transposition can occur during genomic stress events such as polyploidization or hybridization (Gantuz *et al.*, 2022). Lower levels of transpositional activity can be detected during different developmental stages or exposure to environmental stress (Roquis *et al.*, 2021). Integration of TEs at novel locations during transposition is often driven by intrinsic bias for specific sequence or chromatin contexts dependent on TE taxon (Liu *et al.*, 2009; Naito *et al.*, 2009; Quadrana *et al.*, 2019; Roquis *et al.*, 2021). Many TE families preferentially integrate in proximity to protein coding sequences. In such cases the chromatin modifications employed in TE inactivation can influence the sequence accessibility of proximal genes and consequently its expression. In order to limit these occurrences, the Arabidopsis genome has four DNA demethylases, ROS1, DME, DML2 and DML3 to actively reduce methylation at target sequences and maintain TE-gene borders (Erdmann and Picard, 2020). Thus, aside from silencing mechanisms, plant genomes can actively limit repressive chromatin modifications to curb deleterious effects from TE integration.

The capacity of TEs to modify the expression of plant endogenous genes is closely associated with the silencing mechanisms acting on the insert, its location within the genome, and position relative to proximal genes. Within the scope of this review, I will further discuss intrinsic and contextual factors of TEs and their correlation to TE regulation of plant defence genes. Unlike TE-derived structural variation, gene regulation dependent on TEs is not limited to acting on loci containing TE insertions. Thus, I will categorise *cis*- and *trans*-regulatory mechanisms, where *cis*- corresponds to (co-)transcriptional changes of the protein coding sequence with which the TE is associated with and located up/downstream or within the gene itself (Appendix 2). Conversely, I define as *trans*-acting TEs from which ncRNAs are transcribed that can subsequently change the expression of genes not in the vicinity of the source TE sequence (Appendix 2).

3. *Cis* regulation of defence genes by transposable elements

Aside from being a major source of structural variation, transposition can additionally modify the patterns of transcriptional activation of proximal protein coding sequences. (Chuong, Elde and Feschotte, 2017; Hirsch and Springer, 2017; Noshay *et al.*, 2021). Upon integration within novel locations, in most cases the inserted TE will be rapidly inactivated by host's silencing mechanisms (Erdmann and Picard, 2020). In their silenced state TEs can progressively accumulate mutations rendering them permanently immobile (Fambrini *et al.*, 2020). Fixed TEs or TE fragments can become domesticated with new functions in transcriptional regulation or RNA processing of proximal genes at the site of insertion (Fambrini *et al.*, 2020). TEs that are maintained in populations within or in proximity of genes either represent very recent insertions or insertions that are under neutral or positive selection (Fambrini *et al.*, 2020). NLR coding sequences, for example, are significantly enriched for TE insertions, thus implying an adaptive function of this correlation (Quadrana *et al.*, 2016). In continuation, I will further discuss the the mechanisms through

which TEs positively influence the outcome of diverse plant-pathogen interactions by modifying the expression patterns of TE-associated genes (Appendix 3).

3.1. Transposable element insertions in *cis*-regulatory regions

Gene regulatory regions are assemblies of individual transcription binding sites and include promoters, transcriptional enhancers, silencers, and insulator elements (Schmitz, Grotewold and Stam, 2022). Plant defence genes are typically enriched with multiple TF binding sites that are a necessary constituent of defence signal transduction (Schmitz, Grotewold and Stam, 2022). Genome-wide assays of transcriptional activity, open chromatic regions, and TF binding sites have provided compelling evidence that a multitude of TE-derived sequences located within or in proximity of plant regulatory regions can both decrease and increase gene expression (Stuart *et al.*, 2016; Anderson *et al.*, 2019; Uzunović *et al.*, 2019; Wyler *et al.*, 2020; Noshay *et al.*, 2021).Aside from modifying transcription rate of associated genes, TEs can also affect tissue specificity and expression in response to internal and external cues either via co-option of *cis*-regulatory elements or tissue/stress specific chromatin modifications (Hayashi and Yoshida, 2009; Wu *et al.*, 2018; Roquis *et al.*, 2021). These regulatory changes can be derived from regulatory sequences found within TEs or the cellular silencing machinery acting on its locus.

3.1.1 Transposable element derived cis-regulatory elements

Autonomous TEs rely on the host's cell machinery to express the genes required for their transposition (Chuong, Elde and Feschotte, 2017). Consequently, they contain *cis*-regulatory sequences that mimic promoters found in their host genome (Chuong, Elde and Feschotte, 2017; Hermant and Torres-Padilla, 2021). Thus, integration and subsequent domestication of TEs or TE fragments can modify the transcriptional pattern of adjacent genes by providing novel *cis*-regulatory elements (Schmitz, Grotewold and Stam, 2022). The exaptation of TE fragments in *cis*-regulatory modules can be intrinsically determined by TE structural properties. For instance, newly inserted LTR retrotransposons contain two identical flanking sequences at either side of the element's open reading frame (ORF) that control the TEs transcription through promoter-like *cis*-regulatory motifs (Galindo-González *et al.*, 2017). In line with this, LTR retrotransposons are well suited for co-option of *cis*-regulatory elements through domestication of LTR sequences. For example, the 3' LTR of an autonomous *Renovator* element is co-opted as a promoter for *Pit*, a rice resistance gene (Hayashi and Yoshida, 2009). The presence of *Renovator* upstream of the *Pit* sequence constitutively increases its transcription and confers resistance to the rice blast fungus *Magnaporthe oryzae* (Hayashi and Yoshida, 2009). Interestingly, the 5' and 3' LTR of *Renovator* are subject to differential DNA methylation patterns. The 5' LTR contains increased DNA methylation suggesting the transcriptional silencing of TE ORFs. By contrast, the 3' LTR flanking the *Pit* gene has decreased methylation allowing the transcriptional activation of *Pit* through the TE LTR. These results illustrate the specificity of the plant endogenous TE silencing machinery in selectively targeting sequence fragments to inactivate transposition and co-opting adaptive sequences.

Non-autonomous LTR derivatives typically retain the original *cis*-regulatory sequences even after the removal of their endogenous coding regions (Chuong, Elde and Feschotte, 2017; Oka *et al.*, 2017). Conversely, many other non-autonomous elements, such as *LINEs* and *MITEs* truncate the promoter sequence present in the 5' UTR of their ancestral non-LTR TE (Fambrini *et al.*, 2020). A non-LTR TE-derived promoter sequence was recently identified for its function in anthocyanin production in pepper (Jung *et al.*, 2019). The 3' UTR region of the TE contains binding sites for various endogenous TFs that can initiate the transcription of the downstream *CaAn2* anthocyanin regulator. Anthocyanins are plant secondary metabolites that can protect plants from environmental stresses, including pathogen infection (Dong and Lin, 2021). This finding suggests that TE-derived *cis*-regulatory elements do not necessarily originate from the element's regulatory region and therefore may not serve a biological function to the TE itself. Maize husk tissue expression specificity can be conferred by *cis*-regulatory elements found in MITEs, which normally lack regulatory elements driving their own expression (Fagny *et al.*, 2021). Thus, TEs can influence transcription of downstream genes not only through the regulatory elements employed in expression of the TE itself but also regulatory elements functioning independently from TE regulation. Further investigation is required to determine the biological function of these elements within TEs and on proximal endogenous protein coding sequences.

Like canonical host promoters, TE-derived *cis*-regulatory elements often show spatially or temporally regulated activity that is dependent on cell type or environmental cues such as stress or infection (Chuong, Elde and Feschotte, 2017). Activation of TEs by biotic and abiotic stresses has been well characterised in the *Tnt1* family of retrotransposons found in tobacco (Grandbastien *et al.*, 2005). Structural motifs present in the 3' LTR of *Tnt1* initiate its transcriptional reactivation in response to distinct biotic challenges as part of the host's defence pathway (Grandbastien *et al.*, 2005). Recently another family of LTR retrotransposons has been characterised for their sensitivity to biotic stress. The LTR sequences flanking the *ATCOPIA93* coding region were shown to behave as canonical promoters found in immune-defence genes (Zervudacki *et al.*, 2018). Activation of the *ATCOPIA93*-LTR is transiently increased following challenge with bacterial elicitors (Zervudacki *et al.*, 2018). This activation is compromised during infection with virulent strains suggesting that pathogen effectors can attenuate its activation, resembling canonical defence genes. The sensitivity of *ATCOPIA93*-LTR to pathogen molecular patterns is conferred by the presence of W-box regulatory motifs (Zervudacki *et al.*, 2018). W-box elements are cognate binding sites for WRKY TFs which orchestrate the transcriptional reprogramming following initial pathogen recognition (Li *et al.*, 2016). Despite this stress-inducibility, the transcription and consequently the transpositional activity of autonomous *ATCOPIA93* copies in Arabidopsis is limited by DNA methylation in both unchallenged and challenged conditions (Zervudacki *et al.*, 2018). Transcriptional activation of *ATCOPIA93* elements was seen in mutants impaired in DNA methylation but only when treated with defence elicitors (Zervudacki *et al.*, 2018). Thus, the simultaneous induction of defence pathways and de-repression of chromatic silencing is required for *ATCOPIA93* transcription (Figure 1A,B). Transposition intermediates of *ATCOPIA93* were identified in DNA

methylation mutants treated with defence elicitors; however, transposition could not be confirmed (Zervudacki *et al.*, 2018). *ONSEN*, another family of *Copia* TEs, has been previously shown to exhibit similar demethylation-dependent transcriptional activity and transposition during heat stress (Roquis *et al.*, 2021). *ONSEN* transposition induced in laboratory conditions not only conferred heat-stress responsiveness to previously heat-insensitive genes but also induced a myriad of other mutations contributing to variation of genetic sequence and transcriptional regulation (Roquis *et al.*, 2021). Thus, transposition of stress responsive TEs can be a powerful driver of evolution of regulatory networks, but the constitutive presence of repressive chromatin modifications might be necessary to curb the potentially deleterious effects of such transposition events.

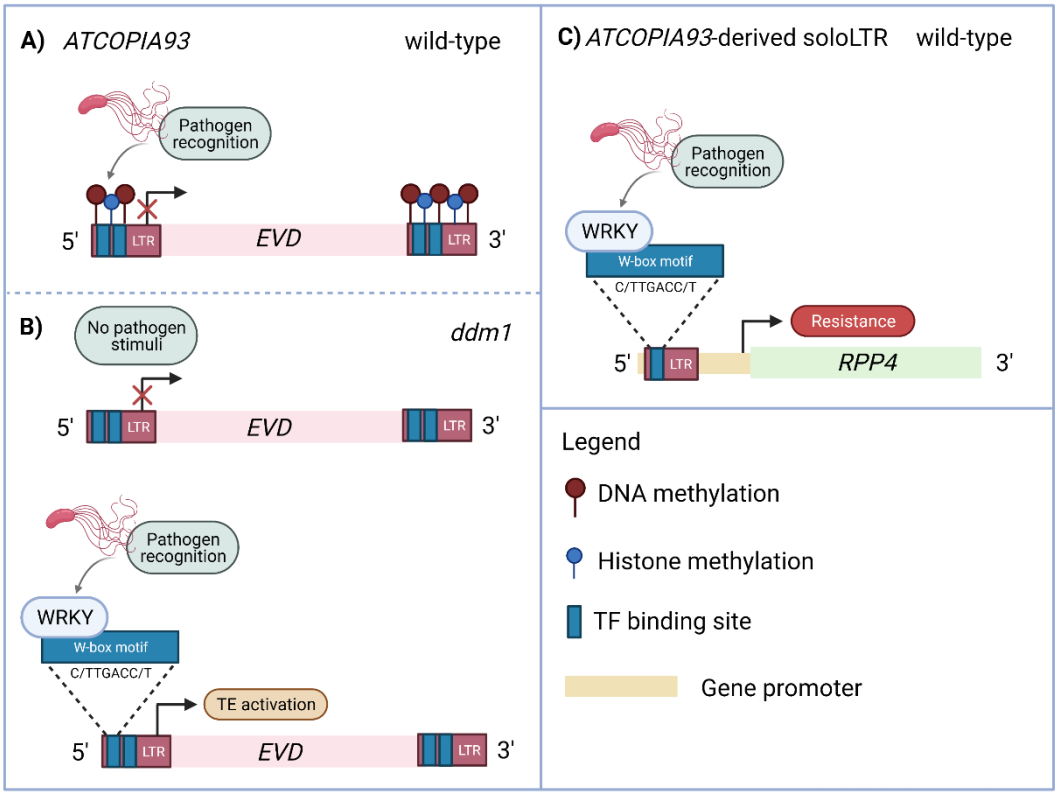

Figure 1. Binding sites (W-box, blue box) for the pathogen responsive WRKY transcription factors are found in the LTR sequences of *ATCOPIA93*. (A) In wild-type plants, the binding of WRKYs to the W-box motif in the autonomous copy of *ATCOPIA93*, *EVD*, is repressed by constitutive silencing by DNA and histone methylation. (B) Plants impaired in the chromatin remodeling factor, DDM1, which functions to suppress TE methylation independently from methylation pathways show sensitivity of EVD1 activation upon pathogen challenge. (C) A non-autonomous copy of a *ATCOPIA93*-derived soloLTR is found upstream of the defence receptor, RPP4. The soloLTR is not repressed by DNA or histone methylation in wild-type plants and actively functions in pathogen signalling pathways. This soloLTR only contains one W-box motif, compared to the found in autonomous copies and mutations in the W-box motif attenuates plant resistance.

Interestingly, an unmethylated soloLTR derived from *ATCOPIA93* was found embedded in the predicted promoter of the *RPP4* gene (Zervudacki *et al.*, 2018). RPP4 is an NLR that confers race-specific resistance against the Arabidopsis downy mildew, *Hyaloperonospora arabidopsidis* but is also responsive to non-specific microbial elicitors. Plants lacking the functional soloLTR or soloLTR-derived W-box in the *RPP4* promoter showed compromised response to an oomycete elicitor (Figure 1C; Zervudacki *et al.*, 2018). Similarly, in Arabidopsis a retro-duplicated LINE, *EPCOT3*, has gained accessible chromatin features typical of *cis*-regulatory sequences, such as methylation of H3K4 (Barco, Kim and Clay, 2019). *EPCOT3* provides a WRKY33 specific binding site and facilitates the re-functionalisation of a novel gene in the conserved pathogen-induced tryptophan biosynthesis pathway (Barco, Kim and Clay, 2019). Thus, TEs that contain *cis*-regulatory elements beneficial to host transcriptional response can undergo modifications of ancestral repressive chromatin features to be co-opted in endogenous signalling. This observation suggests that dynamic demethylation of TEs in response to pathogens is not a necessity for their role in canonical gene regulation. In general, *cis*-regulatory regions derived from older insertions are more likely to be co-opted in gene regulation as they would have had the time to adapt their chromatin modifications (Schmitz, Grotewold and Stam, 2022). Most importantly, the characterisation of *ATCOPIA93* and *EPCOT3* implies that these might not be isolated cases and other TE-derived *cis*-regulatory elements could be found in unrepressed states in service of endogenous defence signalling networks.

3.1.2. Transposable element-dependent (de)methylation of cis-regulatory elements

Active regulatory sequences possess distinct chromatin signatures, characterised with low DNA methylation and active histone modifications (Schmitz, Grotewold and Stam, 2022). TEs found in gene dense regions often have chromatin modifications which are distinct from the local epigenetic state (Fambrini *et al.*, 2020). TE insertions are typically enriched for heterochromatin marks that can repress their transcription (reviewed in section 2). However, these heterochromatic marks are often not restricted to the borders of TE loci but can spread to affect adjacent genes or regulatory elements (Fambrini *et al.*, 2020). Chromatin accessibility of *cis*-regulatory elements in most cases is essential for the initiation and subsequent transcription of protein-coding sequences (Schmitz, Grotewold and Stam, 2022). Despite plant endogenous mechanisms maintaining TE-gene borders, the presence of TE insertions within or in proximity of *cis*-regulatory elements in some cases can disrupt the sequence accessibility and recruitment of cellular transcription machinery (Fambrini *et al.*, 2020).

In plants, the transcriptional activity of stress-responsive genes often correlates with the chromatin states of promoter-associated TEs (Makarevitch *et al.*, 2015; Secco *et al.*, 2015; Espinas, Saze and Saijo, 2016; Fambrini *et al.*, 2020). The chromatic repressive states that silence TEs are reversible and can be actively or passively modified by environmental stressors (Ramakrishnan *et al.*, 2021). In response to pathogen infection, plants can affect the transcriptional regulation of methylation pathway components (Dowen *et al.*, 2012; Yu *et al.*, 2013; López Sánchez *et al.*, 2016; Geng *et al.*, 2019).

Arabidopsis infection with the bacterium *Pseudomonas syringae* DC3000 represses RdDM genes such as *AGO4*, *AGO6*, *NRPD2* and *RDR1* that are necessary for DNA methylation of target loci particularly TEs (Appendix 2; Yu *et al.*, 2013). Expression of NLR genes in Arabidopsis tissues infected with *P. syringae* is accompanied with hypomethylation of NLR-associated TEs (Yu *et al.*, 2013). Similarly, *AGO4* expression is repressed in the wheat progenitor *Aegilops tasuchii* in response to *Blumeria graminis* f. sp. *tritici* (Geng *et al.*, 2019). This reduction in *AGO4* was coupled with a decrease in 24-nt siRNAs biogenesis and global DNA demethylation (Geng *et al.*, 2019). The differentially methylated regions were primarily associated with genes that co-located with TEs and were enriched for annotations in defence function (Geng *et al.*, 2019). Thus, defence-gene associated TEs can provide dynamic transcription in response to pathogens by directing pathogen-dependent chromatin modification at defence loci.

DNA demethylases that function in active reduction of methylation at target loci have likewise been implicated to have a role in plant defence (Le *et al.*, 2014; López Sánchez *et al.*, 2016; Halter *et al.*, 2021; Zeng *et al.*, 2021). In response to *Fusarium oxysporum* infection, the differential expression of over 300 genes is dependent on functional DNA demethylases (Le *et al.*, 2014). A significant proportion of the differentially expressed genes have annotated defence functions, and are enriched for short TE sequences in their promoters (Le *et al.*, 2014). *F. oxysporum* infection does not result in global transcriptional changes of DNA demethylation targets (Le *et al.*, 2014; Schumann *et al.*, 2019), but rather only a specific set of genes are differentially. These findings are consistent with demethylation specificity seen during different biotic stress responses in Arabidopsis (Dowen *et al.*, 2012). Thus, pathogen infection can actively contribute to differential methylation patterns at TE loci associated with defence genes. Interestingly, hyper- and hypomethylation mutants both show increased susceptibility to *F. oxysporum*, and share a significant overlap of differentially expressed genes (Le *et al.*, 2014). This can be attributed to the 'methylstat' in the main DNA demethylase, ROS1 – where methylation of a promoter-associated TE is required for *ROS1* transcription (Lei *et al.*, 2014). Thus, plants impaired in methylation will also be compromised in ROS1 function. In contrast to *F. oxysporum*, infection with *P. syringae* or *H. arabidopsidis* shows a differential response in hypo- and hypermethylated mutants with increased resistance and susceptibility phenotype, respectively (López *et al.*, 2011; Dowen *et al.*, 2012; Yu *et al.*, 2013; López Sánchez *et al.*, 2016)**.** Comparably, the necrotrophic pathogens *Plecetosphaerella cucumerina* and *Alternaria brassicicola* that cause *Plectosporium* blight and black spot disease, respectively, show increased virulence in methylation compromised mutants (López Sánchez *et al.*, 2016). Cumulatively, these findings imply that pathogen-dependent (de)methylation can be specific to different pathogens or at least dependent on the invading pathogen's life cycle. Considering that plant methylation and demethylation patterns largely target TEs it is tempting to speculate that specific sets of TEs can consequentially impact the direct or indirect downstream transcription of defence genes. One question that deserves further investigation is the mechanism of specificity determination during stress dependent chromatin modifications.

The functional mechanisms of how chromatin modifications acting on TEs can affect *cis*-regulatory elements have been illustrated by two remnant *RC/Helitron* TEs found in the promoter of *RMG1* (Figure 2; Halter *et al.*, 2021). *RMG1* is a functional defence gene that requires ROS1-dependent demethylation for an appropriate response to *P. syringae* infection and flagellin treatment (Yu *et al.*, 2013). The ROS1-dependent demethylation of the *RMG1*-associated *Helitron* allows for TF binding to W-box elements found adjacent to the TE sequence (Halter *et al.*, 2021). ROS1 constitutively antagonises RdDM to decrease methylation at the 3' border of the *RMG1* proximal TE insertion (Halter *et al.*, 2021). As expected, the DNA methylation at the *RMG1* TE insertion was almost abolished in *ros1dcl23* mutants which are compromised in the biogenesis of 24-nt siRNAs (Yu *et al.*, 2013). Interestingly, another defence gene regulated by ROS1 does not restore its function in the *ros1dcl23* mutant (Yu *et al.*, 2013). Thus, it seems that ROS1 can regulate specific genes within the same elicitation treatment through different mechanisms. This ROS1-dependent demethylation could function in priming TE-associated defence genes to exacerbate their induction upon pathogen challenge (Halter *et al.*, 2021). The co-option of TE sequences in regulatory modules of stress responsive genes is likely to confer adaptive benefits by providing targets for constitutive or pathogen-induced chromatin modifications to prime or dynamically adapt the intensity of transcription at target genes.

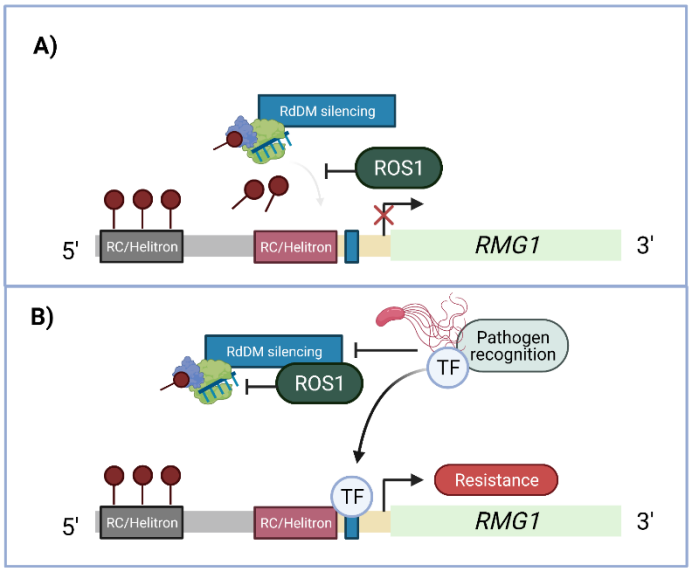

Figure 2. ROS demethylase is required for transcriptional activation of *RMG1* upon pathogen challenge. (A) In unchallenged conditions, ROS1 constitutively inhibits RdDM at proximal *RC/Helitron* TE. The distal *RC/Helitron* insertion is constitutively silenced and is not a target of ROS1. (B) Upon pathogen recognition, various components of the RdDM silencing machinery are downregulated and TFs can bind to the W-box motif found in proximity to the *RC/Helitron* insert.

Aside from DNA methylation, histone modifications can also be dynamically regulated in response to environmental stress. Recently, *Arabidopsis* histone deacetylases and demethylases have been shown to determine the resistance during plant-pathogen interactions (Lee *et al.*, 2016). However, the possible implication of TE insertions in these associations has not yet been reported. Considering the frequency at which histone modifications and particularly repressive marks are associated with TE loci, it can be anticipated that pathogen-dependent histone demethylation likewise targets gene-associated TEs.

3.2. Transposable element insertions within gene body

The effect of TE insertions on intragenic regions is dependent on the location of integration site and the transcriptional activity and functional relevance of the gene (Fambrini *et al.*, 2020). TE insertions into exons or disruptive integration within introns will most often be selected against and may result in the alleles disappearing from the population (Hirsch and Springer, 2017). TEs which integrate within introns are largely tolerated as they are thought to only have a minor impact on the gene's activity. In extreme cases of domestication, the coding sequences of TEs are modified and evolve into a functional role in endogenous signalling. For example, the *Arabidopsis* light-regulated TFs *FHY3* and *FAR1* are derived from the transposase of a *Mutator*-like DNA transposon (Hudson, Lisch and Quail, 2003). More commonly, TEs in genic regions are largely transcriptionally inactive but can be subject to the gene's read through transcription. Thus, the presence of TEs can contribute to the complexity of the plant defence transcriptome by co-transcriptionally altering gene expression (Hirsch and Springer, 2017).

The presence of TEs within introns or exons can lead to alternative splicing of the mRNA transcript potentially altering the protein identity and expression, and therefore its function (Hirsch and Springer, 2017). The exact mechanisms through which TEs can influence the production of splice isoforms is not yet clear and it likely varies in different insertions. Canonically, aberrant splicing is dependent on the presence of intronic/exonic splicing enhancers or silencers (Blake and Lynch, 2021). These sequences can recruit or block RNA-binding proteins which bind to the immature RNA transcript near splice sites to regulate spliceosomal association (Blake and Lynch, 2021). The presence of TEs in the coding region could provision novel regulatory sequences or affect the specificity of existing regulators and thus influence the processing of the mRNA. Additionally, heterochromatic TEs can influence the accessibility of adjacent splice sites and suppress their function by obstructing the association with *trans*-actors such as RNA-binding proteins.

In resistant cucumber genotypes, a non-autonomous retrotransposon insertion in *CsaMLO8* introduces aberrant splicing sites, which results in complete or partial loss of the final exon (Figure 3A; Berg *et al.*, 2015). Plant susceptibility genes can be important determinants of virulence in plant-pathogen interactions (Koseoglou *et al.*, 2022). Infection of powdery mildew, a fungal pathogen, requires expression of *MLO* susceptibility genes in the host plant. Thus, the TE-dependent exon truncation in *CsaMLO8* leads to loss of susceptibility to powdery mildew possibly via incorrect folding of the protein and loss of its function (Figure 3B; Berg *et al.*, 2015). The exact mechanisms through which the TE-insertion in *CsaMLO8* regulates the production of splice isoforms is not yet determined. *MLO* genes are largely conserved across the plant kingdom, and their impairment can confer broad spectrum and durable resistance in plants susceptible to the pathogen. An insertion of a putative transcriptionally active *Ogre* retrotransposon was also found in a resistant allele of *PsMLO1* in pea (Humphry *et al.*, 2011). Like *CsaMLO8*, the insert leads to aberrant splicing and loss of function of *PsMLO1*. However, it seems unlikely that insertional mutations by TE integration is a convergent mechanism conferring loss-of-

function to susceptibility genes. Out of the currently identified natural mutations of *MLO* genes, only *CsaMLO8* and *PsMLO1* contain null mutations attributed to TE insertions (Humphry *et al.*, 2011; Kaufmann *et al.*, 2012; Berg *et al.*, 2015; Nie *et al.*, 2015; Fujimura *et al.*, 2016; Pessina *et al.*, 2016; Lucas *et al.*, 2021). The reason for this is not necessarily the lack of TE integration at other *MLO* loci. For instance, rose *RhMLO3* has a large intronic TE insertion that does not compromise correct gene splicing and is thus unlikely to contribute towards loss of susceptibility (Kaufmann *et al.*, 2012). Arguably smaller scale mutations such as single nucleotide polymorphisms can be just as effective without the potential genome instability implications conferred by TE insertions (Bai *et al.*, 2008; Kaufmann *et al.*, 2012; Fujimura *et al.*, 2016). Therefore, it is likely that TE insertions as a source of inactivating susceptibility genes such as *CsaMLO8* and *PsMLO1* is a redundant mechanism in conferring resistance to pathogens.

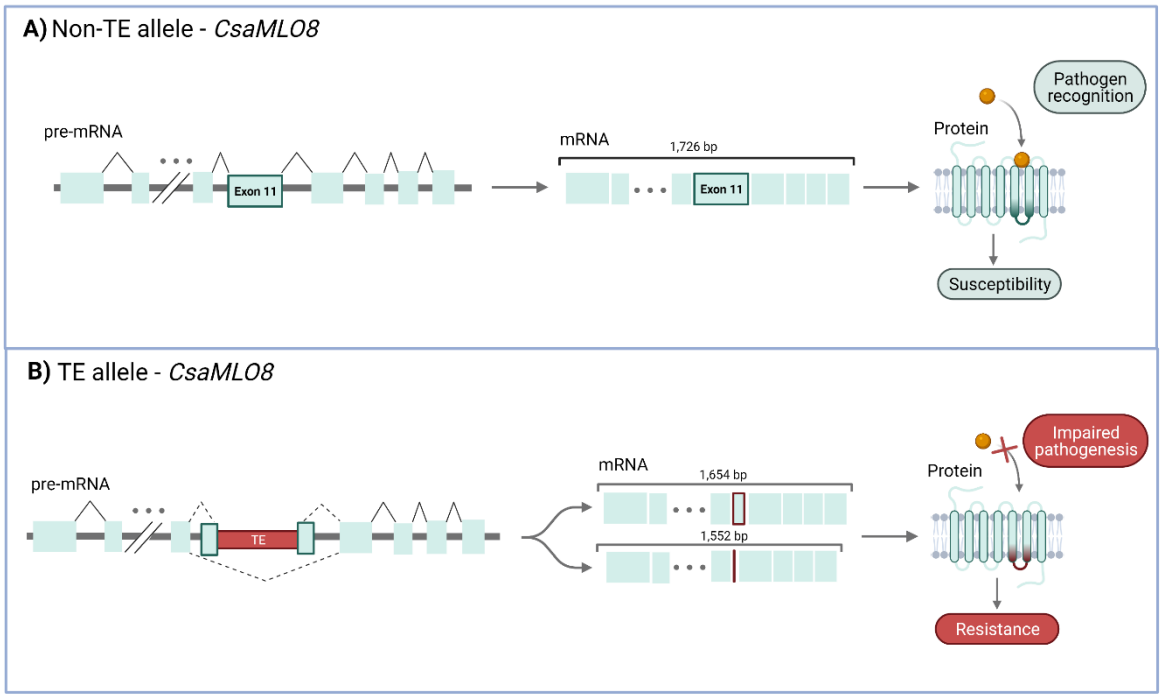

Figure 3. A TE insertion in CsaMLO8 induces aberrant splicing of the final mRNA and consequently compromises the function of the protein product. (A) In CsaMLO8 alleles that do not contain the TE element insert, the transcript is spliced properly producing a functional susceptibility protein. This protein is somehow involved in pathogen infection and confers susceptibility. (B) In plants where the 11th exon of CsaMLO8 contains a TE insert, the pre-mRNA transcript is alternatively spliced producing two transcript isoforms of different lengths, with full or partial deletions of exon 11. The protein product of the aberrant splicing products is predicted to be impaired in an intracellular domain and inhibits the proper function of the protein. Pathogens are unable to utilize this protein in pathogenesis and thus, plants expressing the protein with truncated exon 11 are resistant to powdery mildew.

Aside from alternative splicing, genic TEs can also affect protein function by regulating the polyadenylation of mRNA during its maturation (Hirsch and Springer, 2017). Polyadenylation refers to cleavage of premature transcript at a recognised poly A site and subsequent addition of ~200 adenosine bases to the exposed 3' end, marking the end of the transcript (Blake and Lynch, 2021). The use of

alternative polyadenylation sites (APAS) can alter the functional capacity and stability of the transcript product. TE sequences overlapping with genes can introduce APAS for premature termination of the transcript or limit the usage of sites in their proximity. Alternative polyadenylation is increasingly recognised as a major regulator of plant defence responses, particularly in the expression of functional NLRs (Lai and Eulgem, 2018).

A well-described example is the EDM2 dependent polyadenylation of the TE-associated resistance gene *RPP7* (Tsuchiya and Eulgem, 2013a). Expression of the full-length transcript of *RPP7* is required for Arabidopsis resistance to *H. arabidopsidis* Hiks. An insertion of a *COPIA*-derived retrotransposon (*COPIA-R7*) in the first intron of *RPP7* introduces an APAS found in its 5' LTR (Figure 4; Tsuchiya and Eulgem, 2013a). The repression of the usage of the proximal APAS in *COPIA*-R7 is dependent on the function of EDM2 and two other proteins, which form the ASI1–AIPP1–EDM2 (AAE) complex (You *et al.*, 2021; Zhang *et al.*, 2021). Mutations in the complex elements confers susceptibility of Arabidopsis to *H. arabidopsidis* Hiks accompanied with an increase in transcription of prematurely terminated *RPP7* isoforms (You *et al.*, 2021; Zhang *et al.*, 2021). The AAE complex can recognise and bind to several chromatin marks, such as H3K9me2 which is typically enriched at TE loci including *COPIA-R7* (You *et al.*, 2021). The binding of the AAE complex to *COPIA-R7* consequentially promotes further H3K9 methylation at *COPIA-R7* and represses the usage of the proximal APAS (Tsuchiya and Eulgem, 2013a). Similarly, the 3' untranslated region (UTR) of another race specific NLR, *RPP4*, overlaps with a *Copia* insert with multiple APAS in its 5' LTR (Lai *et al.*, 2020). The correct polyadenylation of *RPP4* and therefore the plant's resistance to *H. arabidopsidis* Emoy2 is dose dependent on EDM2 recruitment of H3K9me2 (Lai *et al.*, 2020). However, EDM2 was found to increase the susceptibility of Arabidopsis in response to *P. syringae* and *H. arabidopsidis* Noco2, suggesting that its function is context specific (Tsuchiya and Eulgem, 2010; Lai *et al.*, 2020). A global analysis of EDM2 function identified that it largely represses the expression of NLR genes and downstream defence components such as WRKY TFs. Thus, RPP7 and RPP4 are an exception in the largely negative association of EDM2 with plant immunity (Lai *et al.*, 2020). Both EDM2-associated and non-associated NLRs are enriched for TE insertions within the gene or in its proximity (Lai *et al.*, 2020). How EDM2 and the AAE complex differentially regulate target TEs is not clear but preliminary research has shown that its dependent on the DNA sequence and existing chromatin modifications (Tsuchiya and Eulgem, 2013b). Interestingly, H3K9me2 at *COPIA-R7* and consequently the expression of the full *RPP7* transcript dynamically decreased in response to *H. arabidopsidis* Hiks infection (Tsuchiya and Eulgem, 2013a). It is unclear whether a transient decrease of NLR expression provides adaptive benefit to plants in mounting pathogen defence response, however this is improbable. Alternatively, pathogen derived effectors can target chromatin modifications at TEs or other loci to promote plant susceptibility. Thus, targeting of TEs by the AAE complex is not only necessary for correct expression of numerous NLRs and downstream defence genes but can also provide a point for dynamic regulation of plant susceptibility. In Arabidopsis, the recurrence of TE-co-option in regulation of alternative polyadenylation in NLR genes is undeniable. In

these cases TEs can provide APAS that promote aberrant transcript termination but also provide further regulatory plasticity as targets for chromatic modifications. The discovery and characterisation of other AAE interacting proteins and AAE-independent pathways in TE-dependent NLR regulation will be instrumental in understanding the complex regulatory networks of NLR expression.

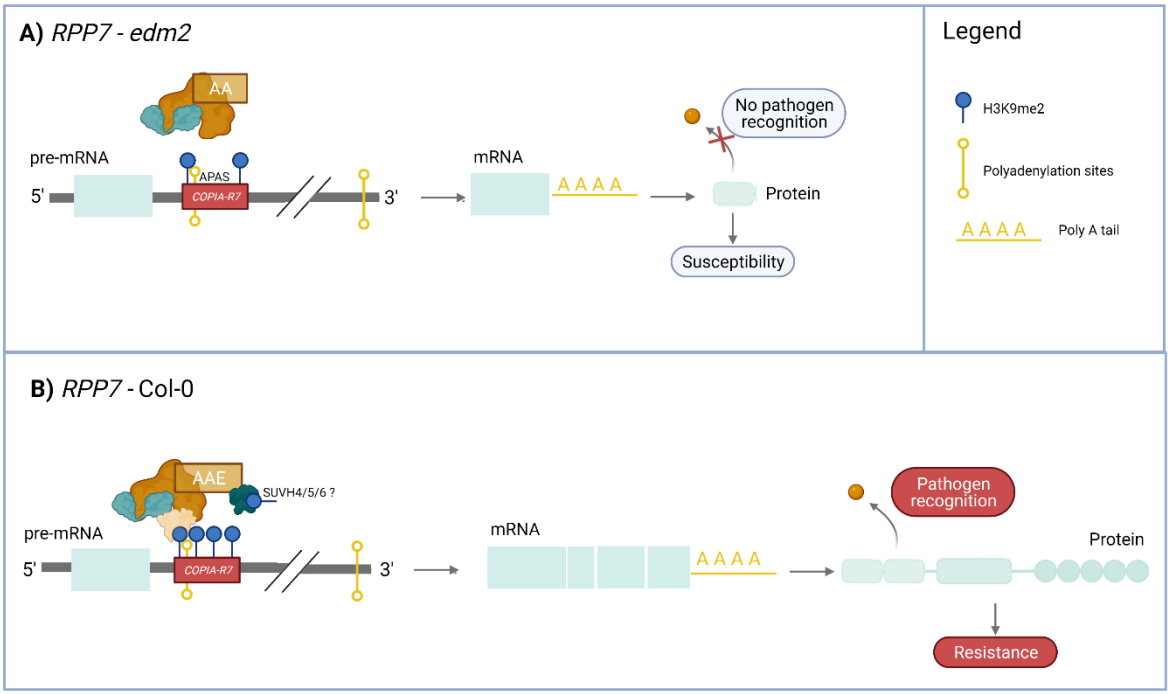

Figure 4. A TE insertion in the first intron of *RPP7* introduces premature APAS and recruits histone modifications through interaction with the AAE complex. (A) In the absence of a functional AAE complex (lacking EDM2), the proximal APAS in *COPIA-R7* is used resulting in the premature termination of the mRNA and a non-functional NLR protein. This protein cannot recognise pathogen effectors and thus, *edm2* mutants show susceptibility to pathogens. (B) In wild-type plants the AAE complex can recognise and bind the histone modifications present at the *COPIA-R7* locus. This interaction promotes further accumulation of H3K9me2 marks at the TE insertion and silences the usage of the TE-derived proximal APAS and promotes the transcription of full length mRNAs. The protein product is a functional NLR that can recognise pathogen effectors and initiates transcription pathways resulting in resistance.

4. Transposable elements as *trans*-actors in regulation of defence genes

The differential expression of defence genes in hypomethylated plants can only be partially explained by TE association in the gene's *cis*-regulatory region (López Sánchez *et al.*, 2016; Halter *et al.*, 2021). Nearly half of the transcriptomic response of Arabidopsis to *H. arabidopsidis* infection is dependent on functional (de)methylation (López Sánchez *et al.*, 2016). However, only 15% of these differentially expressed genes are consistent with DNA methylation changes of genic bodies or proximal TEs (López Sánchez *et al.*, 2016). Alternatively, the maintenance of histone modifications can be dependent on DNA

methylation and therefore impairment of (de)methylation can indirectly affect histone modifications at defence genes influence their transcription. However, the presence of *trans*-regulatory mechanisms is also postulated to drive the expression of defence pathways downstream of pathogen-dependent methylation changes. Multiple models of *trans*-regulation exist by which DNA (de)methylation can regulate defence genes induction. For example, a small number of signalling genes which are directly regulated through a *cis* mechanism can be important regulators in downstream gene expression (Dowen *et al.*, 2012). Therefore, impairing the function of these *cis*-regulated genes will have a global effect on the defence transcriptome. In fact, several (de)methylation dependent defence genes expressed in response to *H. arabidopsidis* included proteins which can activate a wide range of downstream targets (López Sánchez *et al.*, 2016). Other *trans*-actors can also influence the expression of distal TE-associated genes. Here we will consider TE-derived *trans*-actors occurring in the form of ncRNAs that can repress or promote distal gene expression through various RNA-guided mechanisms.

siRNAs that are synthesised for transcriptional silencing of TEs through RdDM can have distal targets at genes with complementary TE insertions (Erdmann and Picard, 2020). For example, a TE insertion within the first intron of *WRKY45* in rice generates TE-siR815 (Zhang *et al.*, 2016), and the presence of the TE and its siRNA product attenuates resistance to *Xanthomonas oryzae* (Zhang *et al.*, 2016). TE-siR815 is complementary to a putative transposon found within *ST1*, an important downstream component in WRKY45-dependent resistance (Figure 5A; Zhang *et al.*, 2016). The methylation levels of *ST1* increase in the presence of TE-siR815 but not in an allelic variant lacking the *WRK45* TE insertion (Zhang *et al.*, 2016). This increase in methylation was compromised in plants that expressed TE-siR815 but were impaired in RdDM function. Thus, TE-siR815 can repress the function of *ST1* as an off-target

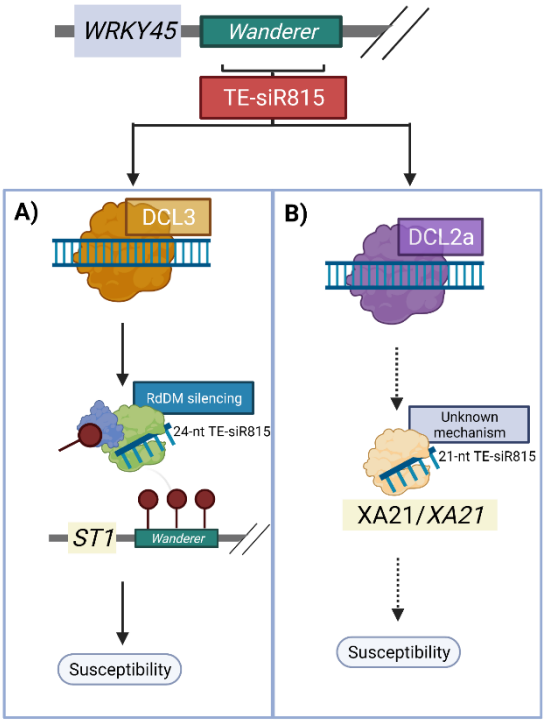

Figure 5. A TE insertion in *WRKY45* is a source of TE-siR815 that can modulate rice resistance to M. oryzae through two independent pathways. (A) TE-siR815 can be processed by DCL3 to produce 24-nt TE-siRNA that is incorporated into AGO4 and the RdDM silencing machinery. RdDM-associated TE-siR815 can target a complementary TE copy found in the intron of *ST1*, a WRKY45 downstream defence gene. Methylation at this locus represses *ST1* expression and induces susceptibility to *M. oryzae*. (B) TE-siR815 can also be processed by DCL2 a to produce 21-nt TE-siRNAs.The 21-nt TE-siRNAs can impair the function of XA21 and promote susceptibility to *M. oryzae*. The pathway through which this TE-siR815 affects defence mediated through XA21 is unknown.

effect dependent on RdDM (Zhang *et al.*, 2016). Interestingly, TE-siR815 was also suggested to confer susceptibility to *X. oryzae* by compromising the function of the immune receptor XA21 in a different rice variety (Figure 5B; Liu *et al.*, 2022). The exact mechanisms through which TE-siR815 inactivates XA21 dependent immunity was however not yet reported. Interestingly however, the susceptibility conferred by TE-siR815 in the two cases are dependent on a different DCL protein for the biogenesis of the siRNA. TE-siR815 biogenesis is dependent on DCL3 and DCL2a in compromising susceptibility through impairment of ST1 and XA21, respectively (Zhang *et al.*, 2016; Liu *et al.*, 2022). Unlike, DCL3, DCL2a is largely involved in maturation of 21-nt siRNAs which are preferentially recruited in PTGS or non-canonical RdDM (Erdmann and Picard, 2020). These results imply that siRNA products from a single TE could have multiple target sites and function through different mechanisms of expression regulation. The ability of TE-derived ncRNAs to differentially target different complementary sequences could provide a powerful mechanism of differentiation and rewiring of transcriptional networks.

In addition to TEs being major sources for endogenous siRNAs, TE sequences have repeatedly been co-opted into coding sequences for microRNA (miRNA) precursors (Hou *et al.*, 2019). Unlike siRNAs that originate from dsRNA synthesised by Pol IV, miRNAs are synthesised by Pol II as single stranded transcripts with secondary hairpin structures (Hou *et al.*, 2019; Erdmann and Picard, 2020). Aside from differences in biogenesis pathways, miRNAs also differ from siRNAs in their function of gene regulation. miRNAs are primarily involved in PTGS by cleaving or inactivating the translation of complementary mRNAs through association with AGO1 (Hou *et al.*, 2019; Erdmann and Picard, 2020).

The role of miRNAs in controlling the defence responses of plants is well recognised. Particularly cereal crops employ many miRNAs to regulate genes involved in susceptibility or defence (Poretti *et al.*, 2020). Around 400 TE-derived miRNAs are differentially expressed in powdery mildew infected wheat tissues and most of these originate from *MITEs*. The non-autonomous *MITEs* are particularly well suited in domestication as miRNA coding genes (Hou *et al.*, 2019). *MITE* sequences are flanked by highly conserved terminal inverted repeats that readily form hairpin structures when expressed. One *MITE*-derived miRNA in particular, Tae_miR1436-1, is transcribed by multiple loci in infected wheat tissues and is consistently expressed only in the presence of pathogens (Poretti *et al.*, 2020). One of the fifty-four complementary sequences identified as potential targets of Tae_miR1436-1, encodes TaeMt3, a protein that participates in an array of protective stress responses (Poretti *et al.*, 2020), some of these responses involves ROS scavenging and regulation of the oxidative burst response associated with cell-death signalling. ROS signalling determines programmed cell-death which is an efficient immune response to biotrophic pathogens, such as powdery mildew (Zhou and Zhang, 2020). *TaeMt3* expression was significantly reduced at 7 days post infection with *B. graminis* which corresponded with a decrease in active Tae-miR1436_1 (Poretti *et al.*, 2020). Thus, Tae_miR1436-1 could target TaeMt3 to possibly contribute to the de-repression of host cell death and in this way contribute towards restriction of the

pathogen. Whether the repression of TaeMt3 and other putative targets by Tae_miR1436-1 occurs transcriptionally or post-transcriptionally is not yet determined.

Despite PTGS being the primary mode of action for miRNA in regulation of gene expression, RdDM can also employ miRNA-derived sequences in DNA methylation of complementary targets (Erdmann and Picard, 2020). miRNAs are canonically sliced by DCL1 to form 21-22-nt transcripts (Li *et al.*, 2019; Erdmann and Picard, 2020). However, at least in rice, DCL3 dependent processing of hairpin miRNA precursors can output 24-nt miRNAs which are compatible for RdDM transcriptional silencing (Li *et al.*, 2019). A recent discovery of a miRNA promoting resistance against *M. oryzae* in rice was able to offer an insight into the biological relevance of this function. miR812w is a Stowaway *MITE*-derived 24-nt miRNA processed by DCL3 (Campo *et al.*, 2021). The expression of miR812w itself is pathogen responsive and transiently activated 48 hours post-inoculation (Figure 6; Campo *et al.*, 2021). Partial deletion of miR812w from a resistant rice cultivar confers susceptibility to *M. oryzae*, whereas overexpression of miR812w transcript increases resistance (Campo *et al.*, 2021). The miR812w dependent resistance is coupled with increased expression of the pathogen resistance marker *PR1* and accumulation of ROS under pathogen challenge (Campo *et al.*, 2021). Almost all (36/39) of the predicted target genes of miR812w contain a complementary *MITE* copy within their UTR. No target cleavage was detected for any of the predicted target sequences (Campo *et al.*, 2021). However, analysis of three putative targets with known

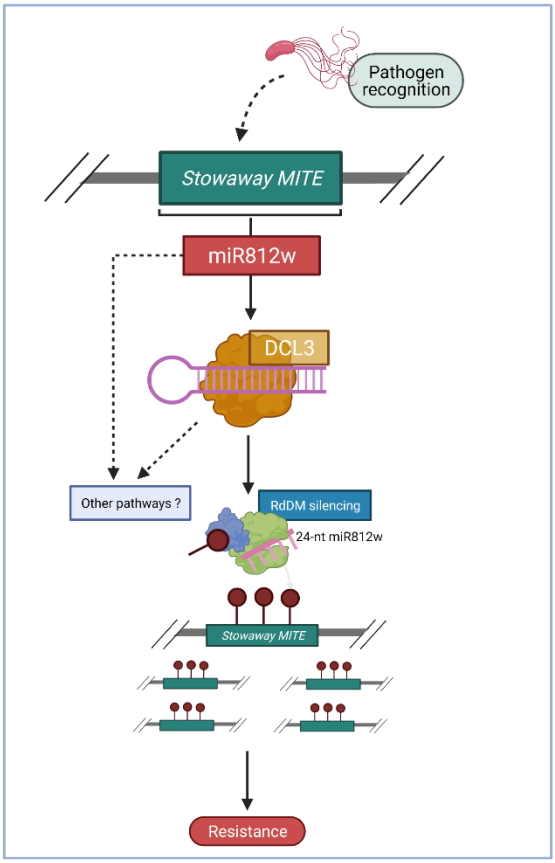

Figure 6. A Stowaway MITE-derived miRNA can be incorporated in the RdDM machinery to transcriptionally expression of target genes. miR812w is synthesized from thus far unknown location in the genome and subsequently processed by DCL3. The 24-nt miRNA product is then loaded onto AGO4 and transcriptionally silences complementary target sequences. Most of the putative targets are annotated defence genes however, the exact genes which are involved in the resistance phenotype are not yet known. Interestingly, miR812w-guided RdDM also targets the MITE insert from where the miRNA originates forming a negative feedback loop. Possibly other pathways of repressing or promoting gene expression exist that incorporate miR812w precursors in their regulation.

annotations as defence gene s show downregulation associated with increased methylation at these loci (Campo *et al.*, 2021). Interestingly, miR812w also methylated its own locus of transcription suggesting it may function in a negative feedback loop (Campo *et al.*, 2021). Thus, TE-derived miRNAs can be used in RdDM at both proximal and distal loci. This could suggest that miRNAs derived from TE-sequences with multiple genic copies can be used for the simultaneous regulation of multiple loci following a single induction event. However, in the case of miR812w, the target loci that contribute towards the resistance phenotype are yet undetermined.

The function of ncRNAs as *trans*-actors is not limited to repression of target gene expression. Recently, a mechanism of *trans*-activation guided by AGO1-bound sRNAs was shown to increase target gene expression by promoting the recruitment of transcriptional machinery (Liu *et al.*, 2018). Canonically, AGO1 functions in the cytoplasm and associates with miRNAs to cleave target mRNAs during PTGS (Erdmann and Picard, 2020). However, nuclear AGO1 guided by 21-nt sRNAs can bind to the chromatin of complementary sequences and consequently increase transcriptional accessibility at target loci (Liu *et al.*, 2018). Treatment of Arabidopsis with different stress elicitors promoted AGO1-dependent binding to treatment-specific set of target genes (Liu *et al.*, 2018). Unlike cytoplasmic AGO1, nuclear AGO1 preferentially associates with *trans*-acting siRNAs (tasiRNAs), which like miRNAs function in PTGS but arise from dsRNA (Liu *et al.*, 2018; Hou *et al.*, 2019). Activated TEs produce RDR6-dependent tasiRNAs-like siRNAs and consequentially these transcripts could be incorporated in various sRNA-guided gene regulatory pathways, including AGO1-dependent transcript upregulation (McCue *et al.*, 2012).Thus, TE-derived sRNAs could promote gene transcription of complementary targets.

In Arabidopsis, infection with *P. syringae* is accompanied with increased biogenesis of siRNAs which show complementarity to TEs and PRR/NLR genes (Cambiagno *et al.*, 2018). Interestingly, at later stages of infection, pericentromeric TEs that associate with the increased siRNA transcripts are silenced, whereas the complementary PRR/NLR genes remain activated (Cambiagno *et al.*, 2018). In unchallenged plants, derepressing the activity of TEs by null-mutations of *MOM1*, a methylation-independent heterochromatin factor that is largely specific to pericentromeric TE silencing, increased the expression of several distal PRR/NLR genes. Plants where repression of pericentromeric TEs is restored only for coregulated MOM1/RdDM loci, showed similar expression of PRR/NLR genes as wild-type (Cambiagno *et al.*, 2018). Therefore, activation of heterochromatic TEs can affect PRR/NLR expression through RdDM-dependent pathways. Whether this function is dependent on RdDM-derived siRNAs guiding AGO1 or an alternative pathway remains to be investigated. Notably, during the analysis of the *mom1* mutants, the PRR/NLR gene with most dramatic transcriptional changes was *RMG1* (Cambiagno *et al.*, 2018). *RMG1* was previously mentioned in this review as ROS1 dependent RdDM reduction at a promoter proximal TE was required for its transcriptional activation during *P. syringae* infection (Halter *et al.*, 2021). Thus, TEs can be directly and indirectly targeted for transcriptional regulation by chromatin modifications and sRNA-guided mechanisms. *RMG1* is not a known target of MOM1 and its H3K9me2 levels remain

unchanged in *mom1* and *P. syringae* infected plants (Cambiagno *et al.*, 2018). It would be interesting to further investigate whether two regulatory mechanisms converge under the same treatment and determine the relation between the *cis*- and *trans*-acting TEs. Particularly, if the distal TE-derived siRNAs are complementary to the *cis*-acting TE insertion in *RMG1* this could indicate the presence of complex regulatory networks dependent on complementary TEs. Demethylation of the *cis*-acting TE could be required for priming of the defence gene loci followed by a transcriptional increase guided by pericentromeric siRNAs therefore enhancing the intensity of *RMG1* expression.

Other than protein coding genes, and the ncRNAs discussed above, long non-coding RNAs (lncRNAs) are a substantial part of the transcriptomic response to pathogen infection in plants (Zhang *et al.*, 2020). lncRNAs can be defined as transcripts of at least 200 bp in size with low protein-coding potential. Diverse mechanisms of gene expression by lncRNAs as activators or repressors of gene expression can finetune and promote the induction of an appropriate defence response (Hou *et al.*, 2019; Zhang *et al.*, 2020). lncRNAs can interact with diverse proteins to control downstream transcription of immune responsive genes, as well as modify alternative splice sites or splicing factors interacting with the pre-mRNA (Seo *et al.*, 2019; Zhang *et al.*, 2020). Post-transcriptionally, lncRNAs can act as decoys for miRNA targets to reduce cleavage of protein-coding mRNA (Jian et al., 2019). Furthermore, lncRNAs can assist gene silencing machineries as precursors for guide sRNAs in transcriptional and post-transcriptional repression, by providing nascent transcript scaffolds during RdDM and recruiting chromatin remodelling proteins (Hou *et al.*, 2019). Many lncRNAs in plants originate from intergenic areas dense with repeats and TEs. These TE-derived lncRNAs have been shown to regulate gene expression for various developmental functions and abiotic stress responses (Wang *et al.*, 2017; Cho, 2018). In tomato, TEs contribute to the lncRNA transcriptional response to biotic stress, however its biological significance was not further investigated (Wang *et al.*, 2017). Considering the commonality of overlap of TE sequences with lncRNAs, as well as the associations between TEs and defence genes it is not unreasonable to assume that TE-derived lncRNAs can function in plant defence. Thus, further research into this question is likely to uncover a new layer of TE-dependent regulation of defence gene expression.

5. Conclusion and importance

The idea that TEs play a fundamental role in the evolution of eukaryotic gene regulation echoes the views first posited by Barbara McClintock during her seminal work on TEs of maize 75 years ago (Fambrini *et al.*, 2020). The research that followed within the next five decades explained TE movement and accumulation as a non-adaptive force, most often having detrimental consequences on the host genome (Dubin, Mittelsten Scheid and Becker, 2018; Fambrini *et al.*, 2020). Recent observations have resurfaced McClintock's original idea and repositioned TEs as causal factors in the evolution of transcriptional gene networks in eukaryotes. Based on the evidence presented here, I would suggest that TEs can directly and indirectly contribute towards the intricacies of the plant defence signalling pathways and that in their absence the evolution and therefore the current function of plant immunity would be impaired.

The structural characteristics and biochemical activities that regulate canonical TE function and mobility to some extent promote the co-option of TE sequences in endogenous plant gene regulation. Most strikingly perhaps, the presence of complete *cis*-regulatory elements in active and inactive TEs could significantly contribute to the diversification and rewiring of defence signalling pathways. Some of these TE regulatory elements show no apparent function in driving TE gene expression or mRNA processing. Nevertheless, these sequences could still be co-opted in regulation of plant endogenous genes following transposon integration. In some cases, however, modification of TE chromatin silencing patterns is required for functional integration in TE sequences n regulatory networks. Silenced TEs can accumulate mutations which can exacerbate *de novo* evolution of regulatory elements but also contribute towards the structural variation of plant genomes. This TE-derived structural variation might be especially important in the diversification of ncRNA sequences and their subsequent integration in finetuning plant defence pathways. Multiple copies of TEs are often found within genomes therefore, this repetition of a TE-derived ncRNA is likely to contribute towards wiring of transcriptional networks and their coordinated response to pathogen infections. The enrichment of TE in immunity and pathogenesis related gene clusters is well characterised for various eukaryotes including plants. NLR genes are often found in clusters enriched with TE insertions that promote duplication of NLRs and their subsequent functional divergence (Quadrana *et al.*, 2016; Lai *et al.*, 2020). We underline another possible aspect of TE-NLR association by TE incorporation in polyadenylation dependent transcription of functional NLRs. The presence of TE-dependent regulation by chromatin modifications might allow plants more flexible expression of defence genes without associated fitness costs. The multiple layers of transcriptional and post-transcriptional control of TEs might additionally provide plants multiple points at which selection can act upon on novel TE insertions to co-opt adaptive sequences whilst silencing TE functional domains to repress transcription. Thus, it is apparent that plants employ TEs not only for sequence variability and evolution of defence genes but also in regulation of these pathways.

The described examples clearly demonstrate the magnitude at which TEs can affect gene expression. TE mobility can contribute towards the evolution and diversification of gene regulatory networks by co-opting of TE-derived *cis*-regulatory elements or redirecting cellular silencing machineries with potential targets in endogenous protein coding genes. In many cases, TE-derived features, such as colour or growth architecture, have been the driving force of artificial selection and consequently the domestication and fixation of the elements promoting these changes. However, understanding the commonality of TE-driven selection for traits with less obvious morphological characteristics, such as pathogen resistance, is lacking. Recently, several QTLs which have been used for introgression in pathogen resistant crop varieties have been shown to owe their phenotypic variability to TE insertions. Perhaps most strikingly, tissue specific expression determined by promoter proximal *MITE* methylation is at the base of a resistance mechanism dependent on two antagonistically acting NLR proteins (Deng *et al.*, 2017). This natural occurring resistance allele has been used consistently for over 30 years without waning of its effectiveness (Deng *et al.*, 2017). Thus, understanding how TEs can modify defence gene

expression is not only essential in ascertaining their adaptive benefits in plant genomes but can also provide powerful tools in innovative techniques for next-generation agriculture. In fact, TEs have already been incorporated in biotechnological procedures for targeted mutagenesis (Yasuda *et al.*, 2013; Ito *et al.*, 2016). The heat sensitive *ONSEN* TE can be potentially used for semi-directional mutagenesis in re-functionalisation of the plant's endogenous genetic material for improved heat-stress response (Ito *et al.*, 2016). Similar procedures can be made for *COPIA93* or additional pathogen sensitive TEs. Additionally, the development of TE-dependent processes for redirection of RNA-guided transcriptional regulatory machinery would provide a powerful tool for both fundamental and applied research. However, whether or not TEs and their transposition is a process that would be safe to incorporate in manipulating crop defence pathways is arguable. The mutagenic effect of TEs remains an undeniable fact of their mobility, thus making the practical application of transposition unpredictable. Therefore, the application of these mechanisms towards agricultural innovations will require much further research to determine their viability.

In future research, it is important to consider the limitations of current studies of TEs. For example, different plant species can have dramatically different TE content, and this can affect the overall role of TE sequences within that genome. Particularly, most current studies of functional characterization of TE gene regulation have been done on the model plant Arabidopsis, whose genome is characterised with a relatively low content of TEs. Thus, translational research is not only necessary to confirm the concepts discussed here but to potentially unearth non-Arabidopsis specific regulatory function of TEs in other genomes.

Our understanding of the plant genomic landscape has increased in complexity with the advent of new technologies and consequent discoveries, leaving us with significantly more questions than we have answers for. In the case of TEs it will be important to determine the adaptive evolutionary benefits and the process of TE domestication in regulatory functions within the plant genomes. TE domestication and their function thereafter is largely dependent on the endogenous cellular silencing machinery. Thus, understanding how TEs evolve differential chromatin modifications from their ancestral copies after transposition will be paramount in determining the course of TE domestication . In answering these questions, it is important to consider the bias of current TE characterisations in regulatory functions. Most transposition events result in detrimental mutations to plant hosts; thus, they are effectively purged from plant genomes. Perhaps the most challenging question to answer will be whether domestication of TEs in regulatory function occur largely by chance, in spite of the mutagenic nature of transpositional events. The alternative answer could imply that plants and TEs have co-evolved mutually beneficial mechanisms that benefits plants in exploiting TE plasticity and allows TEs to persist in plant genomes.

Bibliography

Anderson, S. N. *et al.* (2019) 'Transposable elements contribute to dynamic genome content in maize', *The Plant Journal*, 100(5), pp. 1052–1065. doi: 10.1111/TPJ.14489.

Bai, Y. *et al.* (2008) 'Naturally Occurring Broad-Spectrum Powdery Mildew Resistance in a Central American Tomato Accession Is Caused by Loss of Mlo Function', *Molecular Plant-Microbe Interactions*, 21(1), pp. 30–39. doi: 10.1094/MPMI.

Barco, B., Kim, Y. and Clay, N. K. (2019) 'Expansion of a core regulon by transposable elements promotes Arabidopsis chemical diversity and pathogen defense', *Nature Communications*, 10(1), pp. 1–12. doi: 10.1038/S41467-019-11406-3.

Berg, J. A. *et al.* (2015) 'A transposable element insertion in the susceptibility gene CsaMLO8 results in hypocotyl resistance to powdery mildew in cucumber', *BMC Plant Biology*, 15(1), pp. 1–17. doi: 10.1186/S12870-015-0635-X.

Bhogireddy, S. *et al.* (2021) 'Regulatory non-coding RNAs: a new frontier in regulation of plant biology', *Functional & Integrative Genomics*, 21(3), pp. 313–330. doi: 10.1007/S10142-021-00787-8.

Blake, D. and Lynch, K. W. (2021) 'The three as: Alternative splicing, alternative polyadenylation and their impact on apoptosis in immune function', *Immunological Reviews*, 304(1), pp. 30–50. doi: 10.1111/IMR.13018.

Cambiagno, D. A. *et al.* (2018) 'Immune receptor genes and pericentromeric transposons as targets of common epigenetic regulatory elements', *The Plant Journal*, 96(6), pp. 1178–1190. doi: 10.1111/TPJ.14098.

Campo, S. *et al.* (2021) 'A novel Transposable element-derived microRNA participates in plant immunity to rice blast disease', *Plant biotechnology journal*, 19(9), pp. 1798–1811. doi: 10.1111/PBI.13592.

Cho, J. (2018) 'Transposon-derived non-coding RNAs and their function in plants', *Frontiers in Plant Science*, 9, p. 600. doi: 10.3389/FPLS.2018.00600/BIBTEX.

Chuong, E. B., Elde, N. C. and Feschotte, C. (2017) 'Regulatory activities of transposable elements: from conflicts to benefits', *Nature reviews. Genetics*, 18(2), p. 71. doi: 10.1038/NRG.2016.139.

Deng, Y. *et al.* (2017) 'Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance', *Science*, 355(6328), pp. 962–965. doi: 10.1126/SCIENCE.AAI8898/.

Dong, N. Q. and Lin, H. X. (2021) 'Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions', *Journal of Integrative Plant Biology*, 63(1), pp. 180–209. doi: 10.1111/JIPB.13054.

Dowen, R. H. *et al.* (2012) 'Widespread dynamic DNA methylation in response to biotic stress', *Proceedings of the National Academy of Sciences of the United States of America*, 109(32), p. E2183. doi: 10.1073/PNAS.1209329109/

Dubin, M. J., Mittelsten Scheid, O. and Becker, C. (2018) 'Transposons: a blessing curse', *Current Opinion in Plant Biology*, 42, pp. 23–29. doi: 10.1016/J.PBI.2018.01.003.

Erdmann, R. M. and Picard, C. L. (2020) 'RNA-directed DNA Methylation', *PLOS Genetics*, 16(10), p. e1009034. doi: 10.1371/JOURNAL.PGEN.1009034.

Espinas, N. A., Saze, H. and Saijo, Y. (2016) 'Epigenetic Control of Defense Signaling and Priming in Plants', *Frontiers in plant science*, 7(1201). doi: 10.3389/FPLS.2016.01201.

Fagny, M. *et al.* (2021) 'Identification of Key Tissue-Specific, Biological Processes by Integrating Enhancer Information in Maize Gene Regulatory Networks', *Frontiers in Genetics*, 11, p. 1703. doi: 10.3389/FGENE.2020.606285/BIBTEX.

Fambrini, M. *et al.* (2020) 'The plastic genome: The impact of transposable elements on gene

functionality and genomic structural variations', *Genesis*, 58(12), p. e23399. doi: 10.1002/DVG.23399.

Figueroa-Macías, J. P. *et al.* (2021) 'Plant Growth-Defense Trade-Offs: Molecular Processes Leading to Physiological Changes', *International Journal of Molecular Sciences,* 22(2), p. 693. doi: 10.3390/IJMS22020693.

Fujimura, T. *et al.* (2016) 'Powdery mildew resistance in the Japanese domestic tobacco cultivar Kokubu is associated with aberrant splicing of MLO orthologues', *Plant Pathology*, 65(8), pp. 1358–1365. doi: 10.1111/PPA.12498.

Galindo-González, L. *et al.* (2017) 'LTR-retrotransposons in plants: Engines of evolution', *Gene*, 626, pp. 14–25. doi: 10.1016/J.GENE.2017.04.051.

Geng, S. *et al.* (2019) 'DNA methylation dynamics during the interaction of wheat progenitor Aegilops tauschii with the obligate biotrophic fungus Blumeria graminis f. sp. tritici', *The New Phytologist*, 221(2), p. 1023. doi: 10.1111/NPH.15432.

Gill, R. A. *et al.* (2021) 'On the Role of Transposable Elements in the Regulation of Gene Expression and Subgenomic Interactions in Crop Genomes', 40(2), pp. 157–189. doi: 10.1080/07352689.2021.1920731.

Grandbastien, M. A. *et al.* (2005) 'Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae', *Cytogenetic and Genome Research*, 110(1–4), pp. 229–241. doi: 10.1159/000084957.

Halter, T. *et al.* (2021) 'The Arabidopsis active demethylase ROS1 cis-regulates defence genes by erasing DNA methylation at promoter-regulatory regions', *eLife*, 10, pp. 1–62. doi: 10.7554/ELIFE.62994.

Hannan Parker, A., Wilkinson, S. W. and Ton, J. (2022) 'Epigenetics: a catalyst of plant immunity against pathogens', *New Phytologist*, 233(1), pp. 66–83. doi: 10.1111/NPH.17699.

Hayashi, K. and Yoshida, H. (2009) 'Refunctionalization of the ancient rice blast disease resistance gene Pit by the recruitment of a retrotransposon as a promoter', *The Plant Journal*, 57(3), pp. 413–425. doi: 10.1111/J.1365-313X.2008.03694.X.

Hermant, C. and Torres-Padilla, M. E. (2021) 'TFs for TEs: the transcription factor repertoire of mammalian transposable elements', *Genes & Development*, 35(1–2), p. 22. doi: 10.1101/GAD.344473.120.

Hirsch, C. D. and Springer, N. M. (2017) 'Transposable element influences on gene expression in plants', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1860(1), pp. 157–165. doi: 10.1016/J.BBAGRM.2016.05.010.

Hou, J. *et al.* (2019) 'Non-coding RNAs and transposable elements in plant genomes: emergence, regulatory mechanisms and roles in plant development and stress responses', *Planta 2019 250:1*, 250(1), pp. 23–40. doi: 10.1007/S00425-019-03166-7.

Hudson, M. E., Lisch, D. R. and Quail, P. H. (2003) 'The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway', *The Plant Journal*, 34(4), pp. 453–471. doi: 10.1046/J.1365-313X.2003.01741.X.

Humphry, M. *et al.* (2011) 'Durable broad-spectrum powdery mildew resistance in pea er1 plants is conferred by natural loss-of-function mutations in PsMLO1', *Molecular Plant Pathology*, 12(9), pp. 866–878. doi: 10.1111/J.1364-3703.2011.00718.X.

Ito, H. *et al.* (2016) 'A Stress-Activated Transposon in Arabidopsis Induces Transgenerational Abscisic Acid Insensitivity', *Scientific Reports 2016 6:1*, 6(1), pp. 1–12. doi: 10.1038/SREP23181.

Jung, S. *et al.* (2019) 'A non-LTR retrotransposon activates anthocyanin biosynthesis by regulating a MYB transcription factor in Capsicum annuum', *Plant Science*, 287, p. 110181. doi: 10.1016/J.PLANTSCI.2019.110181.

Kaufmann, H. *et al.* (2012) 'Isolation, molecular characterization, and mapping of four rose MLO

orthologs', *Frontiers in Plant Science*, 3, p. 244. doi: 10.3389/FPLS.2012.00244/BIBTEX.

Koseoglou, E. *et al.* (2022) 'Susceptibility reversed: modified plant susceptibility genes for resistance to bacteria', *Trends in Plant Science*, 27(1), pp. 69–79. doi: 10.1016/J.TPLANTS.2021.07.018.

Lai, Y. *et al.* (2020) 'The Arabidopsis PHD-finger protein EDM2 has multiple roles in balancing NLR immune receptor gene expression', *PLOS Genetics*, 16(9), p. e1008993. doi: 10.1371/JOURNAL.PGEN.1008993.

Lai, Y. and Eulgem, T. (2018) 'Transcript-level expression control of plant NLR genes', *Molecular Plant Pathology*, 19(5), pp. 1267–1281. doi: 10.1111/MPP.12607.

Le, T. N. *et al.* (2014) 'DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis', *Genome Biology*, 15(9). doi: 10.1186/S13059-014-0458-3.

Lee, S. *et al.* (2016) 'Global Regulation of Plant Immunity by Histone Lysine Methyl Transferases', *The Plant Cell*, 28(7), p. 1640. doi: 10.1105/TPC.16.00012.

Lei, M. *et al.* (2014) 'Arabidopsis EDM2 promotes IBM1 distal polyadenylation and regulates genome DNA methylation patterns', *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), pp. 527–532. doi: 10.1073/PNAS.1320106110.

Li, B. *et al.* (2016) 'Transcriptional regulation of pattern-triggered immunity in plants', *Cell host & microbe*, 19(5), p. 641. doi: 10.1016/J.CHOM.2016.04.011.

Li, Y. *et al.* (2019) 'The roles of rice microRNAs in rice-Magnaporthe oryzae interaction', *Phytopathology Research*, 1(1), pp. 1–12. doi: 10.1186/S42483-019-0040-8.

Liu, C. *et al.* (2018) 'Arabidopsis ARGONAUTE 1 Binds Chromatin to Promote Gene Transcription in Response to Hormones and Stresses', *Developmental cell*, 44(3), pp. 348-361. doi: 10.1016/J.DEVCEL.2017.12.002.

Liu, F. *et al.* (2022) 'Silencing of Dicer-like protein 2a restores the resistance phenotype in the rice mutant, sxi4 (suppressor of Xa21-mediated immunity 4)', *The Plant Journal*. doi: 10.1111/TPJ.15692.

Liu, S. *et al.* (2009) 'Mu Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome', *PLoS Genetics*, 5(11), e1000733. doi: 10.1371/JOURNAL.PGEN.1000733.

López, A. *et al.* (2011) 'The RNA Silencing Enzyme RNA Polymerase V Is Required for Plant Immunity', *PLoS Genetics*, 7(12), p. e1002434. doi: 10.1371/JOURNAL.PGEN.1002434.

López Sánchez, A. *et al.* (2016) 'The role of DNA (de)methylation in immune responsiveness of Arabidopsis', *The Plant Journal*, 88(3), p. 361. doi: 10.1111/TPJ.13252.

Lucas, S. J. *et al.* (2021) 'A chromosome-scale genome assembly of European hazel (Corylus avellana L.) reveals targets for crop improvement', *The Plant Journal*, 105(5), pp. 1413–1430. doi: 10.1111/TPJ.15099.

Makarevitch, I. *et al.* (2015) 'Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress', *PLOS Genetics*, 11(1), p. e1004915. doi: 10.1371/JOURNAL.PGEN.1004915.

McCue, A. D. *et al.* (2012) 'Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA', *PLoS Genetics*, 8(2), p. e1002474. doi: 10.1371/JOURNAL.PGEN.1002474.

Naito, K. *et al.* (2009) 'Unexpected consequences of a sudden and massive transposon amplification on rice gene expression', *Nature 2009 461:7267*, 461(7267), pp. 1130–1134. doi: 10.1038/nature08479.

Negi, P., Rai, A. N. and Suprasanna, P. (2016) 'Moving through the stressed genome: Emerging

regulatory roles for transposons in plant stress response', *Frontiers in Plant Science*, 7, p. 1448. doi: 10.3389/FPLS.2016.01448/BIBTEX.

Nie, J. *et al.* (2015) 'Identification and fine mapping of pm5.1: a recessive gene for powdery mildew resistance in cucumber (Cucumis sativus L.)', *Molecular Breeding*, 35(1), pp. 1–13. doi: 10.1007/S11032-015-0206-8/.

Noshay, J. M. *et al.* (2021) 'Assessing the regulatory potential of transposable elements using chromatin accessibility profiles of maize transposons', *Genetics*, 217(1), pp. 1–13. doi: 10.1093/GENETICS/IYAA003.

Noshay, J. M., Crisp, P. A. and Springer, N. M. (2018) 'The Maize Methylome', pp. 81–96. doi: 10.1007/978-3-319-97427-9_6.

Oka, R. *et al.* (2017) 'Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize', *Genome Biology*, 18(1). doi: 10.1186/S13059-017-1273-4.

Pessina, S. *et al.* (2016) 'The knock-down of the expression of MdMLO19 reduces susceptibility to powdery mildew (Podosphaera leucotricha) in apple (Malus domestica)', *Plant Biotechnology Journal*, 14(10), pp. 2033–2044. doi: 10.1111/PBI.12562.

Poretti, M. *et al.* (2020) 'Domestication of High-Copy Transposons Underlays the Wheat Small RNA Response to an Obligate Pathogen', *Molecular Biology and Evolution*, 37(3), pp. 839–848. doi: 10.1093/MOLBEV/MSZ272.

Qiu, Y. and Köhler, C. (2020) 'Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs', *Biochemical Society Transactions*, 48(3), pp. 1005–1017. doi: 10.1042/BST20190937.

Quadrana, L. *et al.* (2016) 'The Arabidopsis thaliana mobilome and its impact at the species level', *eLife*, 5, e15716. doi: 10.7554/ELIFE.15716.

Quadrana, L. *et al.* (2019) 'Transposition favors the generation of large effect mutations that may facilitate rapid adaption', *Nature Communications 2019 10:1*, 10(1), pp. 1–10. doi: 10.1038/S41467-019-11385-5.

Ramakrishnan, M. *et al.* (2021) 'The Dynamism of Transposon Methylation for Plant Development and Stress Adaptation', *International Journal of Molecular Sciences 2021*, 22(21), p. 11387. doi: 10.3390/IJMS222111387.

Roquis, D. *et al.* (2021) 'Genomic impact of stress-induced transposable element mobility in Arabidopsis', *Nucleic Acids Research*, 49(18), pp. 10431–10447. doi: 10.1093/NAR/GKAB828.

Savary, S. *et al.* (2019) 'The global burden of pathogens and pests on major food crops', *Nature Ecology and Evolution*, 3(3), pp. 430–439. doi: 10.1038/S41559-018-0793-Y.

Schmitz, R. J., Grotewold, E. and Stam, M. (2022) 'Cis-regulatory sequences in plants: their importance, discovery, and future challenges', *The Plant Cell*, 34(2), pp. 718–741. doi: 10.1093/PLCELL/KOAB281.

Schumann, U. *et al.* (2019) 'DEMETER plays a role in DNA demethylation and disease response in somatic tissues of Arabidopsis', *Epigenetics*, 14(11), pp. 1074–1087. doi: 10.1080/15592294.2019.1631113.

Secco, D. *et al.* (2015) 'Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements', *eLife*, 4, e09343. doi: 10.7554/ELIFE.09343.001.

Seo, J. S. *et al.* (2019) 'ELF18-INDUCED LONG NONCODING RNA 1 evicts fibrillarin from mediator subunit to enhance PATHOGENESIS-RELATED GENE 1 (PR1) expression', *New Phytologist*, 221(4), pp. 2067–2079. doi: 10.1111/NPH.15530.

Stuart, T. *et al.* (2016) 'Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation', *eLife*, 5, e20777. doi: 10.1101/039511.

Tsuchiya, T. and Eulgem, T. (2010) 'The Arabidopsis defense component EDM2 affects the floral transition in an FLC-dependent manner', *The Plant Journal*, 62(3), pp. 518–528. doi: 10.1111/J.1365-313X.2010.04169.X.

Tsuchiya, T. and Eulgem, T. (2013a) 'An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication', *Proceedings of the National Academy of Sciences of the United States of America*, 110(37), e3535-e3543. doi: 10.1073/PNAS.1312545110/.

Tsuchiya, T. and Eulgem, T. (2013b) 'Mutations in EDM2 selectively affect silencing states of transposons and induce plant developmental plasticity', *Scientific Reports*, 3(1), pp. 1–9. doi: 10.1038/srep01701.

Uzunović, J. *et al.* (2019) 'Transposable Elements Are Important Contributors to Standing Variation in Gene Expression in Capsella Grandiflora', *Molecular Biology and Evolution*, 36(8), pp. 1734–1745. doi: 10.1093/MOLBEV/MSZ098.

Wang, D. *et al.* (2017) 'Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants', *The Plant journal* , 90(1), p. 133. doi: 10.1111/TPJ.13481.

Wu, Q. *et al.* (2018) 'Root-Specific Expression of a Jacalin Lectin Family Protein Gene Requires a Transposable Element Sequence in the Promoter', *Genes*, 9(11), p. 550. doi: 10.3390/GENES9110550.

Wyler, M. *et al.* (2020) 'Impact of Transposable Elements on Methylation and Gene Expression across Natural Accessions of Brachypodium distachyon', *Genome Biology and Evolution*, 12(11), pp. 1994–2001. doi: 10.1093/GBE/EVAA180.

Yasuda, K. *et al.* (2013) 'Utilization of transposable element mPing as a novel genetic tool for modification of the stress response in rice', *Molecular Breeding*, 32(3), pp. 505–516. doi: 10.1007/S11032-013-9885-1/.

You, L. Y. *et al.* (2021) 'Intragenic heterochromatin-mediated alternative polyadenylation modulates miRNA and pollen development in rice', *New Phytologist*, 232(2), pp. 835–852. doi: 10.1111/NPH.17635.

Yu, A. *et al.* (2013) 'Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense', *Proceedings of the National Academy of Sciences of the United States of America*, 110(6), pp. 2389–2394. doi: 10.1073/PNAS.1211757110/.

Zeng, W. *et al.* (2021) 'Roles of DEMETER in regulating DNA methylation in vegetative tissues and pathogen resistance', *Journal of Integrative Plant Biology*, 63(4), pp. 691–706. doi: 10.1111/JIPB.13037/.

Zervudacki, J. *et al.* (2018) 'Transcriptional control and exploitation of an immune-responsive family of plant retrotransposons', *The EMBO Journal*, 37(14), e98481. doi: 10.15252/EMBJ.201798482.

Zhang, H. *et al.* (2016) 'Transposon-derived small RNA is responsible for modified function of WRKY45 locus', *Nature Plants*, 2, e16016. doi: 10.1038/NPLANTS.2016.16.

Zhang, H. *et al.* (2020) 'The Emerging Role of Long Non-Coding RNAs in Plant Defense Against Fungal Stress', *International Journal of Molecular Sciences*, 21(8), 2659. doi: 10.3390/IJMS21082659.

Zhang, X. and Qi, Y. (2019) 'The Landscape of Copia and Gypsy Retrotransposon During Maize Domestication and Improvement', *Frontiers in Plant Science*, 10, p. e1533. doi: 10.3389/FPLS.2019.01533/BIBTEX.

Zhang, Y. Z. *et al.* (2021) 'Genome-wide distribution and functions of the AAE complex in epigenetic regulation in Arabidopsis', *Journal of Integrative Plant Biology*, 63(4), pp. 707–722. doi: 10.1111/JIPB.13068/.

Zhou, J. M. and Zhang, Y. (2020) 'Plant Immunity: Danger Perception and Signaling', *Cell*, 181(5), pp. 978–989. doi: 10.1016/J.CELL.2020.04.028.

Appendix

1. **Literature search strategy**: literature assessed in this review has been obtained from online databases including Scopus and Web of Science. The key words “transposon OR transposable element OR transposition” + “plant” + “defence OR defence OR immunity” + “regulation”
    - From relevant papers cited references and/or citations were tracked using Scite.

2.

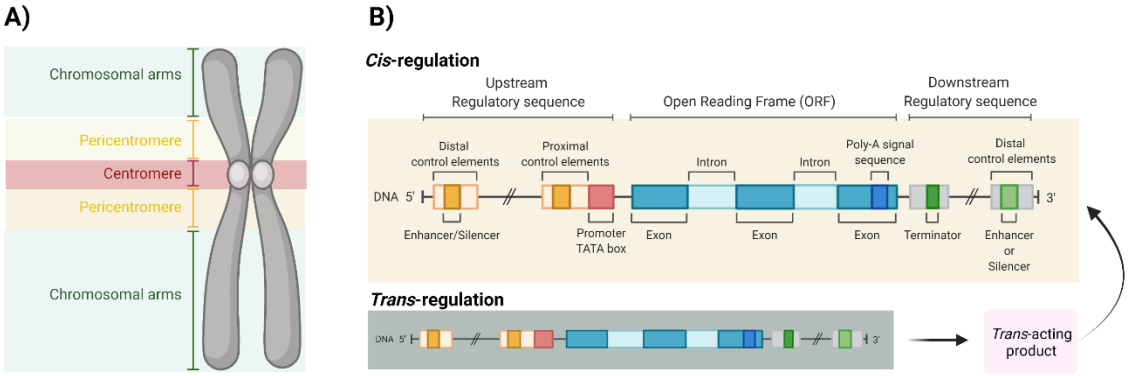

Figure 3. Possible chromosomal positions of TE insertions and types of regulation relative to proximity of affected gene. (A) The centromere is strictly defined as the region of kinetochore formation and spindle attachment during mitosis and meiosis. The chromosomal arms are the chromosomal regions furthest away from the centromere and the pericentromere is found in the middle. The requirement for differential regulation of pericentromeric and TEs in the chromosomal arms stems from the abundance of protein coding sequences in these genomic regions. The pericentromere is generally characterised with repetitive regions derived from TEs and low percentage of functional genes. Thus, constitutive, and more robust heterochromatic structures at these locations is unlikely to affect the expression of endogenous protein-coding sequences. Comparably, repetitive re-establishment by RdDM is required to limit the effect of TE silencing markers acting on protein-coding sequences in the gene dense regions found in chromosomal arms. In Arabidopsis most TEs are found in the pericentromeric regions which is attributed to the strong purifying selection acting in the gene dense chromosomal arms. (B) TEs can affect any of the regulatory sequences or genic components pictured. If the TE insertion is found in any of these elements/regions of the gene with differential expression then we define this as *cis*-regulation. On the other hand, if the TE insertion is found in a gene that produces a trans-acting product that can subsequently affect the transcription of a different gene then this is defines as *trans*-regulation.

3. Supplementary Table 2. A list of known transposable element insertions that dynamically or constitutively modify plants' responses to pathogens.

TE super family	Gene	Region of gene	Plant	Pathogen/elicitor	References
COPIA	*AtRPP4*	Promoter	Arabidopsis	The *COPIA*-soloLTR contains W-box TF binding site for defence responsive WRKYs. Allows RPP4 to respond to oomycete elicitation.	(Zervudacki *et al.*, 2018)
/	*AT5G38550*	Promoter	Arabidopsis	Acts as inhibitor of expression in non-root tissue. Jacalin is a secondary metabolite with a role in plant defence.	(Wu *et al.*, 2018)
/	*CaAn2*	Promoter	Pepper	Promotes accumulation of anthocyanin, a secondary metabolite that has a role in plant defence.	(Jung *et al.*, 2019)
/	*Pit*	Promoter	Rice	Presence of TE constitutively increases expression of Pit gene and confers resistance to *M. oryzae*.	(Hayashi and Yoshida, 2009)
MITE	*PigmS*	Promoter	Rice	TE insertion allows silencing by RdDM in most somatic tissue except for grains where PigmS counteracts the expression of an NLR to rescue growth retardation during defence response.	(Deng *et al.*, 2017)
Helitron	*AtRMG1*	Promoter	Arabidopsis	ROS1 constitutively removes methylation at TE loci proximal to promoter which is necessary for plant defence to pathogens.	(Halter *et al.*, 2021)
LINE	*AtCYP82C2*	Promoter	Arabidopsis	Confers responsiveness to pathogen infection in a duplicated gene resulting in its functional divergence.	(Barco, Kim and Clay, 2019)
/	*CsaMLO8*	Exon	Cucumber	Insertion results in aberrant splicing isoforms with a truncated exon.	(Berg *et al.*, 2015)
Ogre	*PsMLO1*	Intron	Pea	Causes null mutation of susceptibility gene by insertional mutagenesis.	(Humphry *et al.*, 2011)
COPIA	*AtRPP7*	Intron	Arabidopsis	Regulates the access to alternative polyadenylation site found within the TE insertion in the first intron.	(Tsuchiya and Eulgem, 2013)
COPIA	*AtRPP4*	3'UTR	Arabidopsis	Regulates the access to alternative polyadenylation sites found within the TE-gene overlapping 3' region.	(Lai *et al.*, 2020)