An alternative polyadenylation mechanism coopted to the *Arabidopsis RPP7* gene through intronic retrotransposon domestication

Tokuji Tsuchiya and Thomas Eulgem[1]

Center for Plant Cell Biology, Institute for Integrative Genome Biology, Department of Botany and Plant Sciences, University of California, Riverside, CA 92521

Edited by Susan R. Wessler, University of California, Riverside, CA, and approved July 23, 2013 (received for review July 2, 2013)

Transposable elements (TEs) can drive evolution by creating genetic and epigenetic variation. Although examples of adaptive TE insertions are accumulating, proof that epigenetic information carried by such "domesticated" TEs has been coopted to control host gene function is still limited. We show that *COPIA-R7*, a TE inserted into the *Arabidopsis thaliana* disease resistance gene *RPP7* recruited the histone mark H3K9me2 to this locus. H3K9me2 levels at *COPIA-R7* affect the choice between two alternative *RPP7* polyadenylation sites in the pre-mRNA and, thereby, influence the critical balance between RPP7-coding and non–RPP7-coding transcript isoforms. Function of *RPP7* is fully dependent on high levels of H3K9me2 at *COPIA-R7*. We present a direct in vivo demonstration for cooption of a TE-associated histone mark to the epigenetic control of pre-mRNA processing and establish a unique mechanism for regulation of plant immune surveillance gene expression. Our results functionally link a histone mark to alternative polyadenylation and the balance between distinct transcript isoforms from a single gene.

post translational histone modification | EDM2 | *Hyaloperonospora arabidopsidis* | PHD finger

As transposition of transposable elements (TEs) can cause detrimental mutations, TE expression must be tightly suppressed by host silencing mechanisms. In plants, besides methylation of the DNA base cytosine, the posttranslational histone modifications (PHMs) H3K9me2 (dimethylated lysine 9 of histone H3) and H3K27me1 (monomethylated lysine 27 of H3) are closely associated with transcriptional silencing of TEs (1). In *Arabidopsis thaliana* (*Arabidopsis*), H3K9me2 is mainly catalyzed by the partially functionally redundant Su(var)3–9 family histone methyltransferases SUVH4/KRYPTONITE, SUVH5, and SUVH6 (2-5). ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 and 6 (ATXR5 and ATXR6) have overlapping roles in mediating H3K27me1 (6).

Despite their detrimental potenial, TEs can also be beneficial for adaptive evolution of host genome structure and expression control (7–10). For example, TEs have been shown to influence expression of nearby genes by altering local epigenetic states or providing *cis*-regulatory promoter elements. In eukaryotes, the expression of protein-encoding genes is typically regulated at multiple levels including RNA polymerase II (RNAPII)-mediated transcription, pre-mRNA processing, translation, and transcript or protein turnover (11). Alternative polyadenylation (APA) has recently emerged as an important contributor to global gene regulation (12). Differential choice of APA sites (APAS) can affect the protein-coding potential of a given mRNA and/or its stability, localization, or translation efficiency (13).

APA has been thought to be predominantly regulated by polyadenylation factors binding to *cis* elements within pre-mRNAs (14). However, such interactions seem not sufficient to explain all APA-related events observed in vivo. Importantly, RNA processing factors are recruited to pre-mRNA cotranscriptionally when the nascent transcript is still being synthesized at the template genomic DNA (15–17). The physical proximity of the DNA template and transcript-processing events provides opportunities for molecular interactions between chromatin and the pre-mRNA processing machinery. Indeed, recent observations suggested that pre-mRNA processing including polyadenylation and splicing is regulated also at the chromatin level (13, 18, 19). Intriguingly, polyadenylation sites were shown to be depleted of nucleosomes in *Saccharomyces cerevisiae* (20). In addition, differential distribution of PHMs along genes possibly marking distinct architectural features of their sequence has been reported in humans (21).

Plant disease resistance genes encode NLR (nucleotide-binding domain leucine-rich repeat containing)-type immune receptors that trigger defense reactions upon molecular recognition of pathogen-derived molecules (22). We previously reported on the *Arabidopsis* gene *EDM2* (enhanced downy mildew 2), which has a promoting effect on transcript levels of the NLR gene resistance to *Peronospora parasitica* (*RPP7*) (23). Although we also showed EDM2 to contribute to transcriptional TE silencing by modulating levels of the repressive PHM H3K9me2 (24), it has been unclear how EDM2 affects *RPP7*. Here, we show that EDM2 affects levels of RPP7-coding transcripts by modulating APA. We further demonstrate that this APA mechanism results from cooption of epigenetic information at a TE insertion locus. Besides mechanistic insight on chromatin-level control of APA, we provide an example for biologically relevant effects of TE insertions on gene function.

Results

***EDM2* Controls *COPIA-R7*–Associated H3K9me2 and Co- or Post-transcriptionally Affects RPP7-Coding Transcript Levels.** NLR expression and activity is known to be tightly controlled, as certain minimal receptor activity levels are required for efficient pathogen

Significance

We show the histone mark H3K9me2, which is known to mediate transposon silencing, to determine the choice between alternative polyadenylation sites within the *Arabidopsis thaliana* disease resistance gene *RPP7*. High H3K9me2 levels recruited to the first *RPP7* intron by the *COPIA-R7* retrotransposon suppress use of a promoter-proximal polyadenylation site. Modulating H3K9me2 levels at this site shifts the balance between full-length and incomplete RPP7 transcripts. By recruiting H3K9me2-dependent polyadenylation control to *RPP7*, the *COPIA-R7* insertion provided a new switch to fine-tune *RPP7* expression. Selective advantages resulting from this mechanism likely contributed to the domestication of *COPIA-R7* at *RPP7*.

Author contributions: T.T. and T.E. designed research; T.T. performed research; T.T. analyzed data; and T.T. and T.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Database deposition: The nucleic acid sequences reported in this paper have been deposited in the GenBank database (accession nos. KF112064–KF112069).

See Commentary on page 14821.

[1]To whom correspondence should be addressed. E-mail: thomas.eulgem@ucr.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1312545110/-/DCSupplemental.

recognition whereas NLR activity levels above a certain threshold can trigger autoimmunity and spontaneous cell death (25–29). We found levels of RPP7-coding transcripts to be well-correlated with levels of immunity conferred by this disease resistance gene, indicating that proper control of these transcripts is critical for *RPP7* function (30) (Fig. S1).

RPP7 (AT1G58602) in the *Arabidopsis* Col-0 accession is a complex gene with three noncoding exons upstream from its start codon, followed by three coding exons and three noncoding exons (Fig. 1*A*). The first *RPP7* intron contains a *Ty-1 COPIA*-type retrotransposon in sense orientation that we termed *COPIA-R7* (Fig. 1*A*). *COPIA-R7* has long terminal repeats (LTRs) at both ends.

By RT-PCR, we previously found levels of RPP7-coding transcripts to be strongly reduced in *edm2* mutants compared with their parental wild-type background Col-0 (23). We confirmed the reduction of spliced RPP7-coding transcript levels in three independent *edm2* mutants by real-time quantitative (q) RT-PCR (primer combination c in Fig. 1*B*). By chromatin immunoprecipitation (ChIP) coupled with qPCR, we found H3K9me2 levels in *COPIA-R7* to be high in Col-0 and strongly reduced in *edm2*

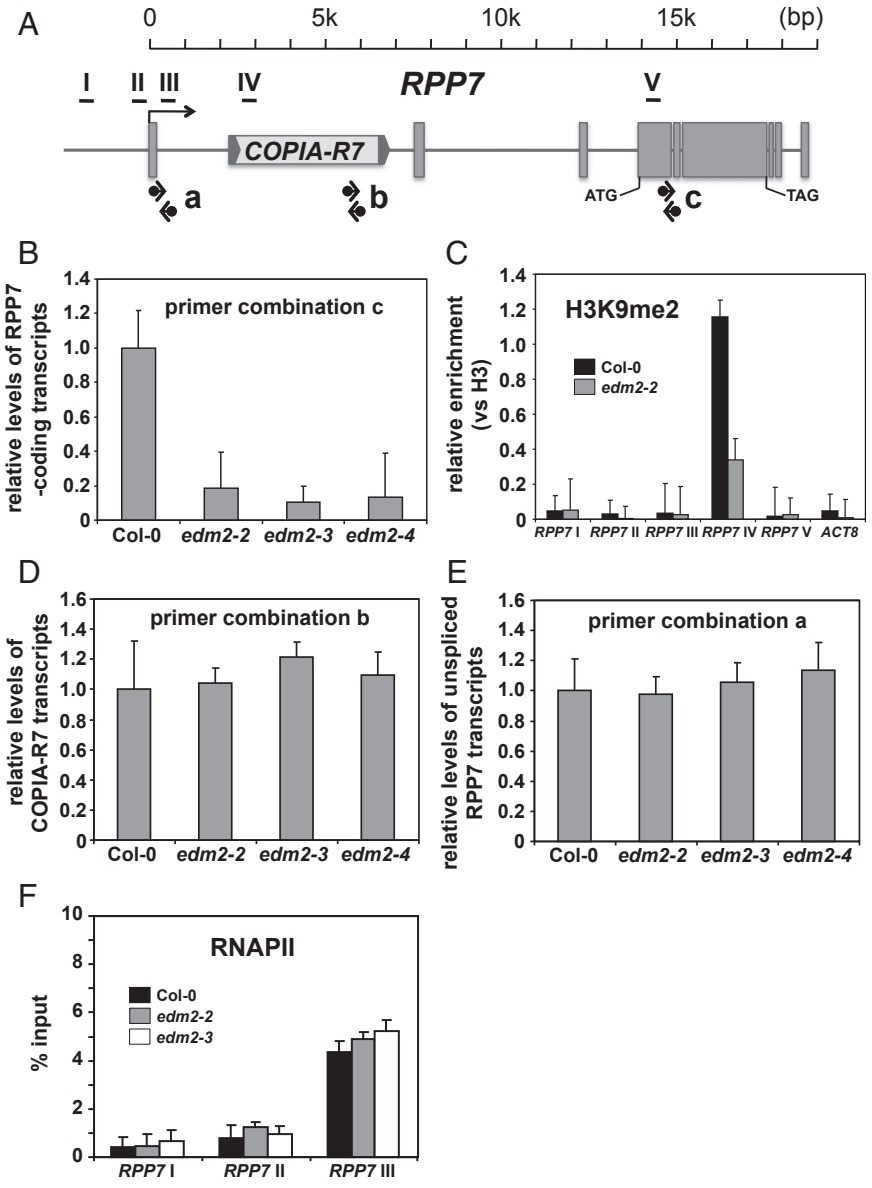

Fig. 1. Effects of EDM2 on transcripts and H3K9me2 at the *RPP7* locus. (*A*) Schematic representation of the *RPP7* locus in the *Arabidopsis* Col-0 accession. *RPP7* exons are shown as gray-filled boxes. Two LTRs flanking the body of *COPIA-R7* are represented as gray-filled arrowheads. Horizontal bars (I, II, III, IV, and V) denote regions analyzed by ChIP shown in *B* and *F*. Black arrows indicate positions of PCR primers used for analyses shown in *B*, *D*, and *E*. (*B*) Levels of RPP7-coding transcripts (spliced mature transcripts) measured by qRT-PCR using primer combination c annealing to sequences in exon4 and exon5. Error bars represent SD for three independent experiments. (*C*) ChIP-qPCR to measure H3K9me2 levels at *RPP7*/*COPIA-R7*. The *y* axis represents H3K9me2 levels normalized to histone H3 occupancy. *ACTIN8* (*ACT8*) serves as a control locus. Error bars represent SEM for two biological replicates with three technical replicates each. (*D*) COPIA-R7 transcript levels determined by qRT-PCR with primer combination b annealing to sequences within the transposon. Error bars represent SD for three independent experiments. (*E*) Levels of nascent (unspliced) RPP7 transcripts determined by qRT-PCR with primer combination a annealing to sequences in exon1 and intron1. Error bars represent SD for three independent experiments. (*F*) ChIP-qPCR to measure RNAPII occupancy at three *RPP7* regions surrounding the transcription start site. The *y* axis represents RNAPII levels normalized to total input. Error bars represent SEM for two biological replicates with three technical replicates each.

mutants (Fig. 1*C*). Levels of this histone mark were low and not influenced by *EDM2* in regions of *RPP7* remote from the TE. Levels of other well-characterized PHMs, such as H3K4me3, H3K27me1, and H3K27me3 are not or only slightly affected at *RPP7*/*COPIA-R7* in *edm2* mutants (Fig. S2). Despite the reduction of H3K9me2 levels at *COPIA-R7,* qRT-PCR analysis showed transcript levels of this TE not to be altered in *edm2* plants (Fig. 1*D*).

To understand whether reduced RPP7-coding transcript levels in *edm2* mutants can be accounted for by a change of the rate of transcription, we measured unspliced pre-mRNA transcripts in a population of total RNA by qRT-PCR (primer combination a in Fig. 1 *A* and *E*). This PCR technique of nascent RNA determination was successfully used previously (31). Unlike mature RPP7-coding mRNA, levels of nascent (unspliced) RPP7 transcripts are not reduced in *edm2* mutants. Consistently, RNAPII occupancy at *RPP7* regions surrounding the transcription start site, as determined by ChIP-qPCR, was also not significantly altered in *edm2* mutants (Fig. 1*F*). These data demonstrate that *EDM2* promotes high H3K9me2 levels at *COPIA-R7* and positively affects levels of RPP7-coding transcripts in a co- or posttranscriptional manner.

EDM2 Affects the Ratio Between Two Distinct RPP7 RNA Transcript Isoforms. We separately measured by qRT-PCR levels of transcripts containing each *RPP7* exon (exon1–exon4; Fig. 2*A*, *Upper*).

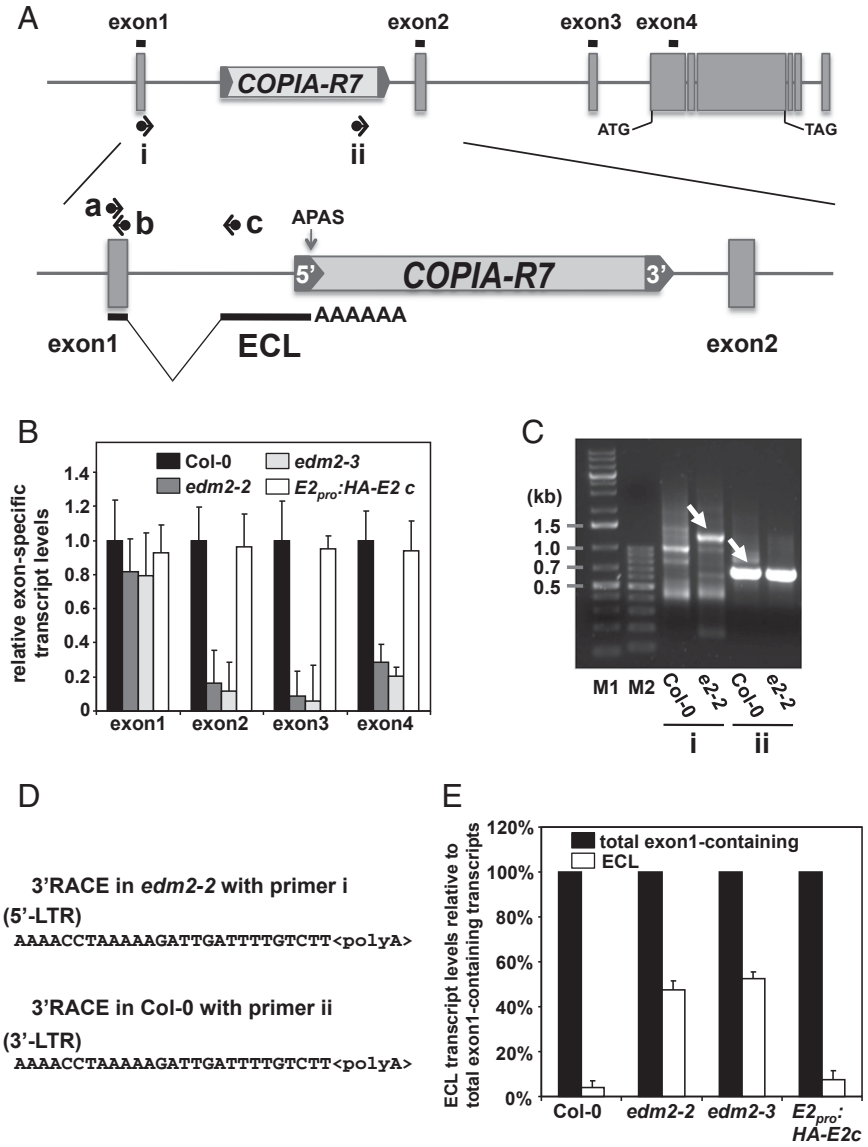

Fig. 2. Effects of *EDM2* on the ratio of *RPP7* RNA transcript isoforms. (*A*) Schematic representation of *RPP7* with the RNA transcript isoform ECL. Black horizontal lines above each exon represent regions amplified by qRT-PCR in the experiment shown in *B*. Black horizontal arrows represent PCR primers used for 3′RACE (i and ii) or qRT-PCR (a–c) in *C* or *E*, respectively. The alternative polyadenylation site (APAS) preferentially used in *edm2* mutants is marked by a vertical arrow. (*B*) Levels of exon-specific transcripts measured by qRT-PCR. Error bars represent SD for three independent experiments. (*C*) The 3′RACE analysis of poly(A)-selected total mRNAs extracted from Col-0 or *edm2-2* mutant plants with primers i and ii represented in the *Upper* section of *A*. The bands marked with white arrows were further analyzed by sequencing. M1, GeneRuler 1 kb Plus DNA ladder (Fermentas). M2, GeneRuler 100 bp DNA ladder (Fermentas). (*D*) Polyadenylation sites of ECL (*Upper*) or COPIA-R7 (*Lower*) revealed by 3′RACE experiments. (*E*) Ratios of ECL transcripts relative to total exon1-containing transcripts determined by qRT-PCR. Total exon1-containing transcripts were determined using primers a and b (*A*) whereas ECL transcripts were measured using primers a and c (*A*). Error bars represent SD for three independent experiments.

Levels of transcripts containing exon1 of *RPP7* did not differ between Col-0 and *edm2* plants whereas levels of transcripts containing the later *RPP7* exons are clearly reduced in these mutants (Fig. 2*B*). This effect was reverted to the wild-type Col-0 level in a transgenic complementation line expressing N-terminally HA-tagged EDM2 driven by the native *EDM2* promoter in the *edm2-2* background ($E2_{pro}$:*HA-E2c*). This line also exhibited fully restored *RPP7*-mediated immunity and wild-type levels of RPP7-coding transcripts (Fig. S1). These observations strongly suggested that an EDM2-dependent regulatory mechanism targets intron1 of *RPP7*.

We further analyzed RPP7 transcripts by 3′RACE (rapid amplification of cDNA ends) with RNA from Col-0 and *edm2-2* plants (Fig. 2*A*, primers i and ii). Similar band patterns were observed between 3′RACE runs with Col-0 and *edm2-2* RNA using primers ii, after agarose gel electrophoresis. However, band patterns obtained with primer i clearly differed between Col-0 and *edm2-2* (Fig. 2*C*). By sequencing a differential 3′RACE product generated from *edm2-2* RNA with primer i (indicated by an arrow in Fig. 2*C*), we identified one spliced RPP7 transcript species that is initiated at the primer annealing site in exon1, extends until a regular splice donor site at the end of the exon1, and continues from an alternative splice acceptor site (ASAS) within intron1 until it is terminated in the 5′-LTR of *COPIA-R7* (Fig. 2*A*, *Lower*). We termed this shorter non–RPP7-coding transcript ECL (exon 1-containing LTR-terminated transcript). Sequencing the 3′RACE product for Col-0 with primer ii (Fig. 2*C*, highlighted by an arrow) revealed the polyadenylation site of COPIA-R7 transcripts to be located in the 3′-LTR of this TE. The 5′-LTRs of retrotransposons often act as promoters whereas 3′-LTRs function as terminators although their nucleotide sequences of 5′- and 3′-LTRs are near-identical (32) (Fig. S3). The region harboring the polyadenylation site for ECL in the 5′-LTR is identical to that for COPIA-R7 in the 3′-LTR (Fig. 2*D*).

By 5′RACE, we found a cluster of common transcription start sites (TSSs) for ECL and RPP7-coding transcripts spreading over a very short genomic stretch (Fig. S4) surrounding the predicted *RPP7* TSS (www.arabidopsis.org/). Thus, both types of transcripts must be controlled by the same promoter. ECL is unlikely to encode any functional protein as the polypeptide potentially encoded by the longest ECL ORF consists of only 129 amino acids, does not start from an ATG codon, and does not have obvious homology (E > 1.8) to any protein from *Arabidopsis* or other organisms in protein databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Taken together, these data revealed the presence of alternative polyadenylation (APAS) and splice acceptor sites (ASAS) in intron1 of *RPP7*.

By qRT-PCR, we found the fraction of ECL transcripts within the pool of exon1-containing transcripts to be substantially higher in *edm2-2* plants compared with Col-0 (Fig. 2 *A* and *E*). The balance between these two transcript types was reverted to that observed in Col-0 in the $E2_{pro}$:*HA-EDM2c* line (Fig. 2*E*). Together with the fact that the rate of de novo RPP7 transcript synthesis is not affected by *EDM2* mutations, this observation shows the APAS to be more strongly used in *edm2* mutants compared with Col-0. In *edm2* plants, preferential use of this promoter-proximal APAS results in enhanced levels of non–RPP7-coding ECL transcripts among all exon1-containing transcripts and, consequently, reduced levels of RPP7-coding transcripts.

EDM2 Physically Associates with EDM2-Dependent H3K9me2-Marked Areas at *ECL*/*COPIA-R7*. A high-resolution analysis of H3K9me2 levels at *RPP7* by ChIP-qPCR showed high levels of this mark to be present from exon2 of *ECL* until exon2 of *RPP7* including the body of *COPIA-R7* (Fig. 3 *A* and *B*, regions 4–10). *RPP7* regions upstream from the splice acceptor site of *ECL* exon2 exhibit only low levels of H3K9me2 (Fig. 3*B*, regions 1–3). Intriguingly, the extent of H3K9me2 in *ECL* exon2 and *COPIA-R7* is clearly dependent on *EDM2*. In this region, levels of this PHM are substantially lower in *edm2* plants compared with Col-0 whereas near wild-type levels are restored in the $E2_{pro}$:*HA-EDM2c* line.

By ChIP-qPCR with the $E2_{pro}$:*HA-EDM2c* line and a commercially available anti-HA antibody, we observed clear HA-EDM2 enrichment at regions 5–10 (Fig. 3*C*), indicating that EDM2 is physically recruited to an area, stretching from *ECL* exon2 into *COPIA-R7,* that carries EDM2-dependent H3K9me2 marks.

A Triple Mutant of H3K9 Methyltransferases Phenocopies *edm2* Mutants. EDM2 affects H3K9me2 levels at *ECL*/*COPIA-R7* and balances the ratio between ECL and RPP7-coding transcripts. If the effect of EDM2 on H3K9me2 is causal for the ECL/RPP7 coding transcript balance, the *suvh456* triple mutant, which is deficient in all three major *Arabidopsis* H3K9 methyltransferases, should phenocopy all *RPP7*-related phenotypes of *edm2* mutants. As expected, the *suvh456* mutant exhibited strongly reduced H3K9me2 levels in exon2 of *ECL* and *COPIA-R7* compared with Col-0 (Fig. 4*A*). We also observed in this triple mutant reduced levels of RPP7-coding transcripts (Fig. 4*B*), enhanced ECL transcript levels (Fig. 4*C*), and complete loss of *RPP7*-dependent *Hyaloperonospora arabidopsidis* resistance (Fig. 4*E*). As seen in *edm2* mutants, reduction of H3K9me2 levels in *suvh456* did not affect the rate of *RPP7* transcription (Fig. 4*D*). Thus, reduced H3K9me2 levels are indeed causal for the altered balance between ECL and RPP7-coding transcripts and the loss-of-immunity phenotype of *edm2* mutants.

Insertion of *COPIA-R7* Coopted the H3K9me2-Dependent Regulatory Mechanism to *RPP7* Expression Control. Inspection of the genomes of 80 wild inbred *Arabidopsis* accessions (http://1001genomes.org/index.html) (33) indicated that most natural variants of this species contain *COPIA-R7* in *RPP7*. However, six *Arabidopsis* accessions—Krazo-2, Koch-1, Cdm-1, Istisu-1, ICE75, and ICE134—appeared to lack this TE in intron1 of *RPP7*. Sequencing of cloned genomic sequences from Krazo-2 and Koch-1 confirmed the lack of *COPIA-R7* in the respective regions. Based on cDNAs we cloned from Krazo-2 and Koch-1, the exon/intron structures of their *RPP7*-like genes and Col-0 *RPP7* are clearly conserved. However, *COPIA-R7* and fragments of other transposon in intron2 (Fig. S5) are lacking in the RPP7-like genes from Krazo-2 and Koch-1 (Fig. 5*A*). Because of their high similarity to Col-0 *RPP7* both at the nucleotide and deduced amino acid sequence levels, we concluded that these genes are orthologs of Col-0 *RPP7* and were therefore named $RPP7_{Krazo\text{-}2}$ or $RPP7_{Koch\text{-}1}$, respectively (Fig. 5*A*). Detailed sequence analyses of $RPP7_{Krazo\text{-}2}$ or $RPP7_{Koch\text{-}1}$ are described in *SI Text*.

Intriguingly, we found insertional sequence blocks at the position corresponding to the stretch immediately downstream to Col-0 *COPIA-R7* in both $RPP7_{Krazo\text{-}2}$ and $RPP7_{Koch\text{-}1}$. This sequence is neither related to transposons nor homologous to any Col-0 genomic sequences. We term it NIC (not in Col-0) (Fig. 5*A*). Surprisingly, by 3′RACE (Fig. S6), we found ECL-like transcripts also to be generated at $RPP7_{Krazo\text{-}2}$ and $RPP7_{Koch\text{-}1}$. Sequencing of the 3′RACE products revealed that the ASAS for ECL-like transcripts in $RPP7_{Krazo\text{-}2}$ and $RPP7_{Koch\text{-}1}$ are conserved compared with that in Col-0 *RPP7* (Fig. 5*B*) whereas their APAS are located in the middle of the NIC sequences (Fig. 5*B*).

We cloned cDNAs from Krazo-2 and Koch-1, which are orthologous to that of EDM2 in Col-0. At the nucleotide level, they are 99.4% and 100% identical to the Col-0 EDM2 cDNA, respectively. Their deduced amino acid sequences are also 99.4% and 100% identical to that of Col-0 EDM2. Col-0, Krazo-2, and Koch-1 plants expressing a common EDM2 mRNA silencing trigger exhibited strongly reduced levels of EDM2-like transcripts (Fig. 5*C*). These lines showed morphological changes of leaves also seen in *edm2* mutants (34). Therefore, the cloned Krazo-2 and Koch-1 cDNAs clearly encode EDM2 orthologs.

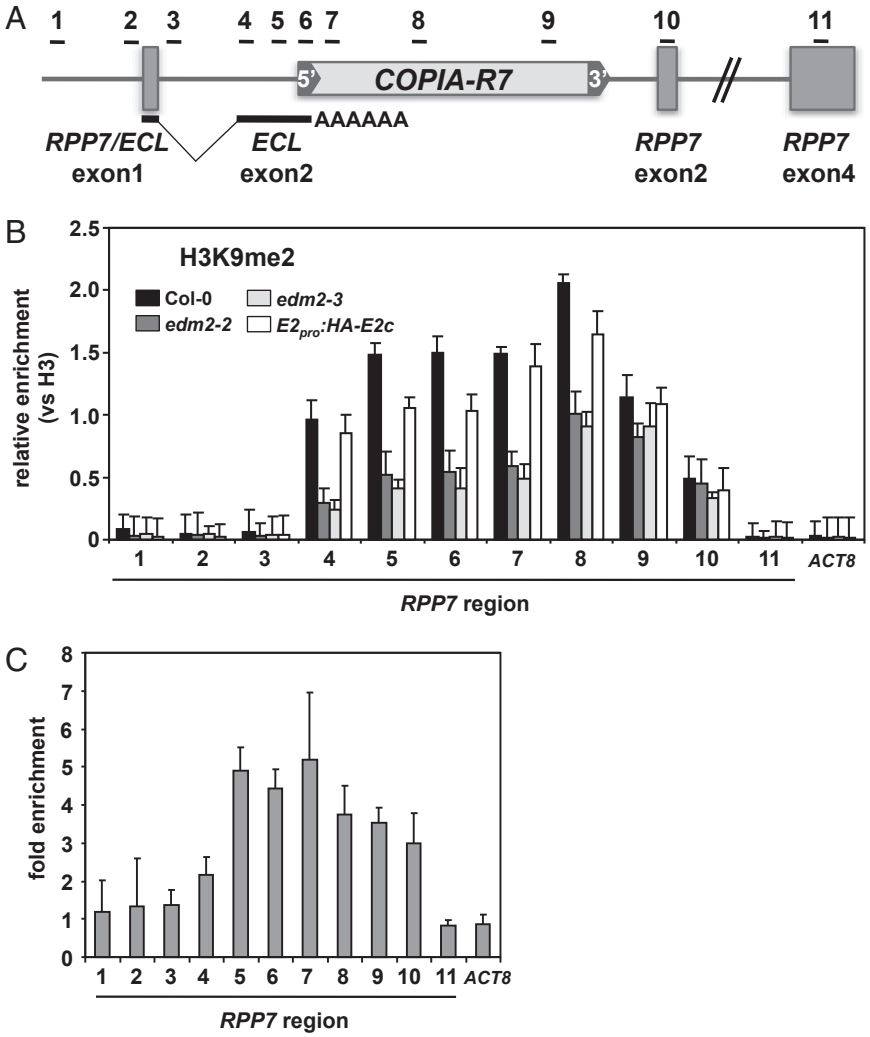

Fig. 3. EDM2 is physically associated with H3K9me2-marked chromatin regions at *RPP7/COPIA-R7*. (*A*) Schematic representation of the upstream region of the *RPP7* locus. Horizontal bars represent areas examined by ChIP-qPCR in *B* and *C*. (*B*) H3K9me2 levels determined by ChIP-qPCR in Col-0, *edm2-2*, *edm2-3*, and the $E2_{pro}$*:HA-E2c* complementation line. The *y* axis represents H3K9me2 levels normalized to histone H3 occupancy. *ACTIN8* (*ACT8*) served as a control locus. Error bars represent SEM for two biological replicates with three technical replicates each. (*C*) Levels of HA-tagged EDM2 (HA-EDM2) at *RPP7* in $E2_{pro}$*:HA-E2c* plants. ChIP-qPCR was pefromed with an anti-HA tag antibody. *ACTIN8* (*ACT8*) served as a control locus. The *y* axis represents fold enrichment of signals in the $E2_{pro}$*:HA-E2c* line relative to those in Col-0. Error bars represent SEM for two biological replicates with three technical replicates each.

Therefore, we termed the respective genes $EDM2_{Krazo\text{-}2}$ and $EDM2_{Koch\text{-}1}$.

Silencing of $EDM2_{Krazo\text{-}2}$ and $EDM2_{Koch\text{-}1}$ did not affect levels of RPP7-coding transcripts whereas silencing of EDM2 in Col-0 clearly reduced levels of these transcripts (Fig. 5*D*). In addition, silencing of $EDM2_{Krazo\text{-}2}$ and $EDM2_{Koch\text{-}1}$ also did not affect levels of $ECL_{Krazo\text{-}2}$ and $ECL_{Koch\text{-}1}$ (Fig. 5*E*). Therefore, *EDM2* does not control *RPP7* expression in these accessions.

We also measured levels of H3K9me2 at $ECL_{Krazo\text{-}2}$, $ECL_{Koch\text{-}1}$, and Col-0 *ECL*, respectively (Fig. 5*F*). Consistent with the data in Fig. 3*B*, high levels of H3K9me2 were detected in exon2 of *ECL* in Col-0 (regions 2–3 in Fig. 5*B*). This PHM is depleted at the corresponding *ECL* regions of Krazo-2 and Koch-1 (regions ii and iii in Fig. 5*B*). Thus, high levels of H3K9me2 at *ECL* in Col-0 must have originated from the *COPIA-R7* insertion and *EDM2* affects *RPP7* expression through *COPIA-R7* in Col-0.

H3K9me2-Dependent APA Fine-Tunes RPP7-Coding Transcript Levels in Response to *Hpa*Hiks1. We previously reported a transient up-regulation of RPP7-coding transcripts following *Hpa*Hiks1 recognition in Col-0 (23). By qRT-PCR analysis in Col-0, we now found this *Hpa*Hiks1-induced transient increase of RPP7-coding transcript levels to be accompanied by up-regulation of all exon1-containing transcripts including those of ECL (Fig. 6*A*). Most importantly, measuring relative levels of RPP7-coding transcript levels compared with those of ECL (R7/ECL), we found the R7/ECL ratio to decline after *Hpa*Hiks recognition (Fig. 6*B*). This trend was tightly correlated with a similar *Hpa*Hiks1-induced decline of H3K9me2 at the APAS for ECL (Fig. 6*C*). Such pathogen recognition-induced changes of the R7/ECL ratio, H3K9me2 levels at the APAS for ECL, or RPP7-coding transcript levels were not observed in *edm2-2* (Fig. 6 *B* and *C*) (23). Multiple mechanisms are likely to adjust levels of RPP7 coding transcripts after *Hpa*Hiks1 infection. However, based on our observations, dynamic changes of the R7/ECL balance caused by modulation of H3K9me2 at the APAS for ECL clearly contribute to the fine-tuning of RPP7 coding transcript levels.

Discussion

Here, we report a unique mechanism controlling expression of the *Arabidopsis* disease resistance gene *RPP7* that involves

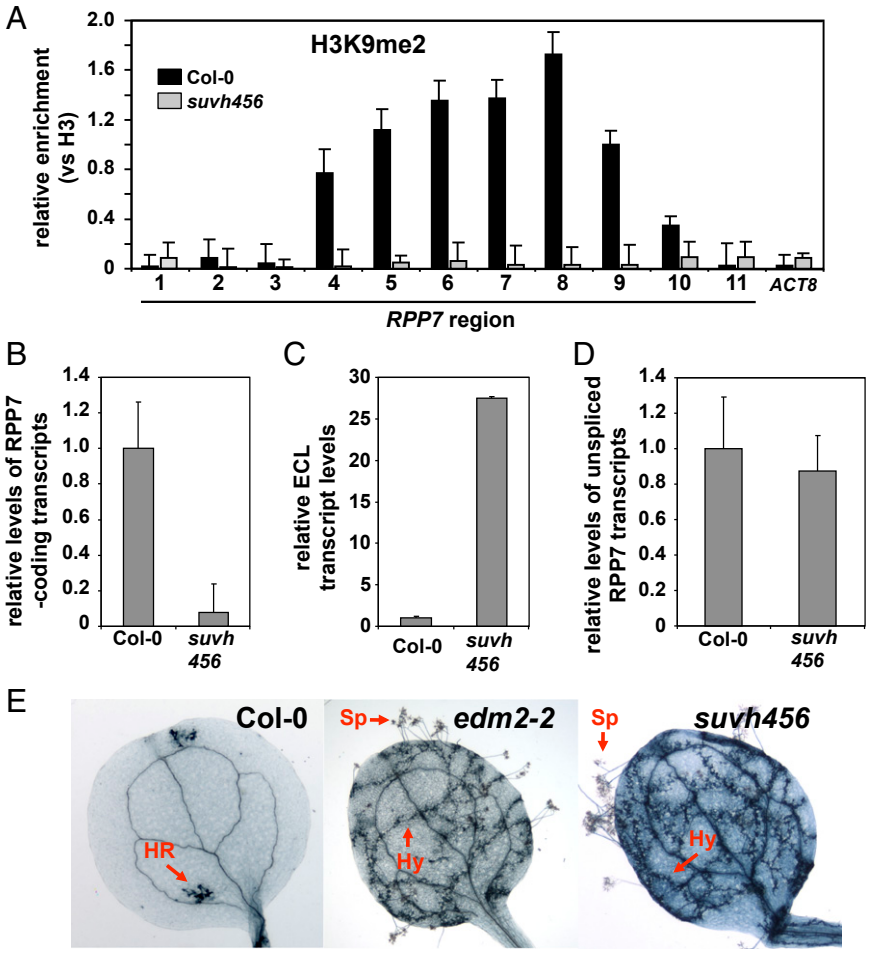

Fig. 4. The Col-0 *suvh456* triple mutant phenocopies *RPP7*-related effects of *edm2* mutants. (*A*) H3K9me2 levels determined by ChIP-qPCR in Col-0, *edm2-2*, and *suvh456*. The *y* axis represents H3K9me2 levels normalized to histone H3 occupancy. *ACTIN8* (*ACT8*) served as a control locus. Error bars represent SEM for two biological replicates with three technical replicates each. (*B*) Levels of RPP7-coding transcripts (spliced mature transcripts) measured by qRT-PCR using primer combination b (Fig. 1*A*) located in exon4 and exon5 of *RPP7* in Col-0, *edm2-2*, *edm2-3*, and *suvh456*. Error bars represent SD for three independent experiments. (*C*) Levels of ECL transcripts measured by qRT-PCR using primers a and c (Fig. 2*A*) in Col-0, *edm2-2*, *edm2-3*, and *suvh456*. Error bars represent SD for three independent experiments. (*D*) Levels of nascent (unspliced) *RPP7* transcripts determined by qRT-PCR with primer combination a (Fig. 1*A*) annealing to sequences in exon1 and intron1 of *RPP7*. Error bars represent SD for three independent experiments. (*E*) Typical trypan blue-stained cotyledons of 2-wk-old Col-0, *edm2-2*, and *suvh456* seedlings 7 d postinfection with 5×10^4 spores of the *H. arabidopsidis* (*Hpa*) isolate Hiks1 mL^{-1}. This *Hpa* isolate is highly specifically recognized by RPP7 (23, 41).Trypan blue stains *Hpa* structures as well as plant hypersensitive response (HR) sites dark blue. HR, HR site; Hy, *Hpa* hyphae; Sp, *Hpa* sporangiophores. HR is a programmed cell death response closely associated with immunity against *Hpa* whereas Hy and Sp occur when immunity fails and the plant is susceptible to *Hpa* (23, 30).

H3K9me2-mediated APA and determines the balance between two distinct *RPP7*-derived transcript isoforms (Fig. 7). This mechanism is a consequence of the insertion of the retrotransposon *COPIA-R7* in intron1 of *RPP7* and is critical for the biological function of *RPP7*. The plant homeodomain (PHD)-finger protein EDM2 that we previously showed to control H3K9me2 and silencing states of the *Arabidopsis* TEs *Mu1* and *COPIA4* (24) also affects this histone mark at *COPIA-R7* and has been coopted along with *COPIA-R7* to the control of *RPP7* expression.

COPIA-R7 has two key roles in this unique mechanism. Firstly, the 5′-LTR of this TE provides an APAS. Secondly, this retrotransposon carries epigenetic information that controls use of this APAS. Lower levels of H3K9me2 at the *COPIA-R7* region in *edm2* mutant plants correlate with increased use of this promoter-proximal APAS in the 5′-LTR of *COPIA-R7*, resulting in enhanced levels of ECL transcript and reduced levels of RPP7-coding transcript.

The physical association of EDM2 with the genomic region of *ECL* exon2 as well as *COPIA-R7*, where high levels of EDM2-dependent H3K9me2 are observed, indicates direct participation of this protein in this gene regulatory mechanism. It is well-established in *Arabidopsis* that the repressive PHM H3K9me2 predominantly associates with TEs and contributes to heterochromatic silencing (1). Furthermore, H3K9me2 appears to have spread from the body of *COPIA-R7* into *RPP7* intron1.

Mutations of *EDM2* do not affect the rate of *RPP7* transcription. The degree of immunity conferred by this disease resistance gene is positively correlated with levels of the RPP7 RNA isoform, which encodes an immune receptor. Therefore, a regulatory mechanism in *RPP7* expression through APA can act as a switch to define levels of the biologically functional mRNA. Indeed, we found this gene regulatory mechanism to fine-tune RPP7-coding transcript levels in responses to pathogen.

In the *suvh456* triple mutant, H3K9me2 is depleted at all of the tested *RPP7* regions, indicating that at least one of the SUVH4, -5, and -6 methyltransferases is responsible for this histone mark at *RPP7*. Similar to *edm2* mutants, we observed, in *suvh456* plants, enhanced use of the promoter-proximal APAS, resulting in

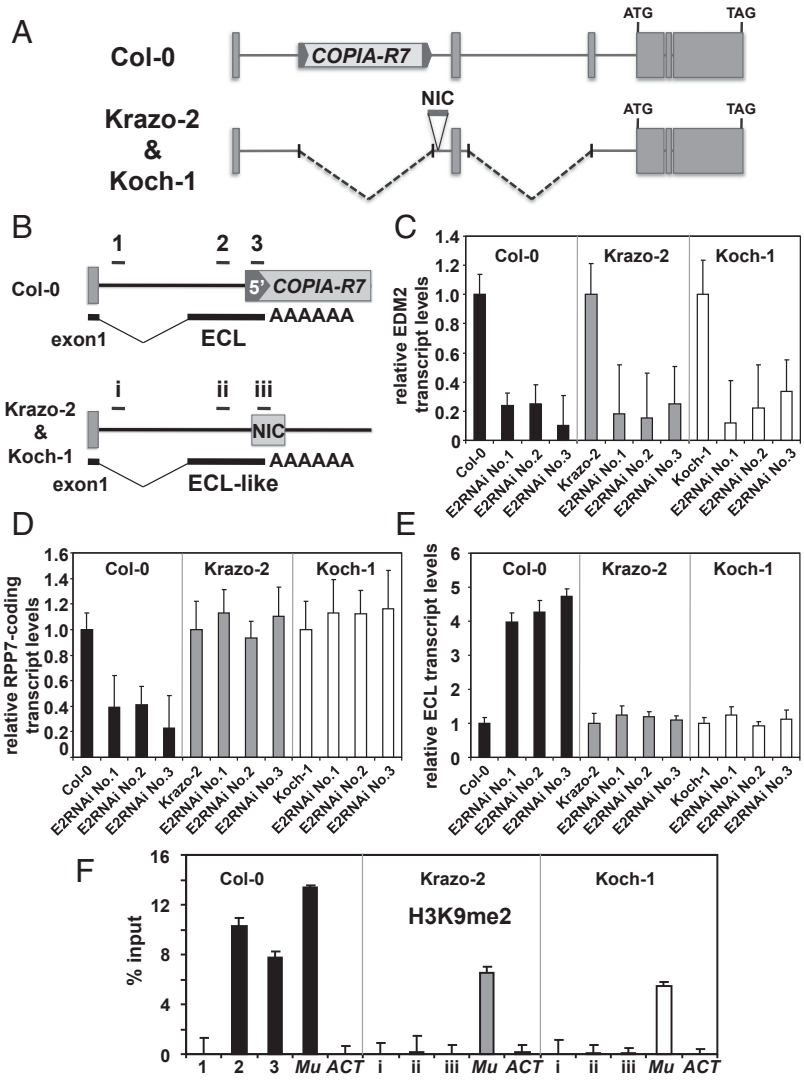

Fig. 5. Effects of *EDM2* on *RPP7* expression in *Arabidopsis* natural accessions. (*A*) Schematic representation of the *RPP7* locus in *Arabidopsis* Col-0, Krazo-2, and Koch-1 accessions. *RPP7* exons are shown as gray-filled boxes. The genomic DNA sequence block NIC in Krazo-2/Koch-1 is represented as gray-filled horizontal bars. (*B*) Schematic representation of the upstream region of *RPP7* locus with the RNA transcript isoform *ECL* in Col-0 and Krazo-2/Koch-1. Horizontal bars (1–3 and i–iii) denote regions analyzed by ChIP for H3K9me2 in *F*. (*C*) Transcript levels of *EDM2* determined by qRT-PCR in Col-0, Krazo-2, Koch-1, and *EDM2* silencing lines in the background of each of these three *Arabidopsis* accessions. Error bars represent SD for three independent experiments. (*D*) Levels of RPP7-coding transcripts determined by qRT-PCR in Col-0, Krazo-2, Koch-1, and the respective EDM2 silencing lines. Error bars represent SD for three independent experiments. (*E*) Levels of ECL transcripts determined by qRT-PCR in Col-0, Krazo-2, Koch-1, and the respective EDM2 silencing lines. Error bars represent SD for three independent experiments. (*F*) ChIP-qPCR to measure H3K9me2 levels at *RPP7* in Col-0, Krazo-2, and Koch-1. The *y* axis represents H3K9me2 levels normalized to input DNA. The *ECL* regions represented in *B* were tested. The DNA transposon *Mu1* (*Mu*) and *Actin8* (*ACT*) served as positive and negative controls, respectively. Error bars represent SEM for two biological replicates with three technical replicates each.

enhanced ECL transcript and reduced RPP7-coding transcript levels. These observations clearly show that H3K9me2 can trigger this gene regulation mechanism and confirms the effects of EDM2 on *RPP7* expression to be driven by changes in H3K9me2 levels.

In addition to histone modifications, nucleosome positioning has been implicated as a determinant of the choice of alternative polyadenylation sites in humans (19). However, we did not observe any significant difference of total histone H3 levels at any of the tested *RPP7* regions in *edm2* mutants compared with those in Col-0 (Fig. S7). Thus, enhanced use of the promoter-proximal APAS in *edm2* mutants possibly is solely the consequence of reduced levels of H3K9me2 and not due to changes in nucleosome density.

Only a small number of natural *A. thaliana* variants seem to lack *COPIA-R7* in their *RPP7* orthologs. We experimentally confirmed the absence of *COPIA-R7* in the two *Arabidopsis* accessions Krazo-2 and Koch-1. Besides *COPIA-R7*, short fragments derived from different types of TEs are present in intron1 and intron2 of Col-0 *RPP7* (Fig. S5). Their presence is most likely the consequence of repeated TE insertions and their removal by illegitimate recombination (35). Thus, the *COPIA-R7* insertion in intron1 is likely beneficial for *RPP7* and, therefore, got stabilized whereas other TE insertions in intron1 and intron2 were not domesticated.

In Krazo-2 and Koch-1, expression control of *RPP7* was shown to be independent of *EDM2*, indicating that the *COPIA-R7* insertion in *RPP7* is required for the influence of *EDM2* on *RPP7* expression. Surprisingly, we found ECL transcripts also to be generated in Krazo-2 and Koch-1 although no TE-related DNA sequences exist in intron1 of their respective *RPP7* orthologs. Instead, a sequence block that is not present in Col-0 *RPP7* was found to act as the terminator of ECL transcripts in Krazo-2 and

Koch-1. Thus, Krazo-2 and Koch-1 also use a mechanism of APA-mediated balance between ECL and full RPP7-coding transcripts. In these accessions, however, this mechanism is not under EDM2-mediated H3K9me2 control. Based on our observations, it is plausible that the APA-mediated *RPP7* expression mechanism was developed in ancestral *A. thaliana* plants before Col-0 and Krazo-2/Koch-1 evolutionally diverged. Thus, the insertion of *COPIA-R7* carrying epigenetic information of H3K9me2 into *RPP7* intron1 of Col-0 and utilization of this transposon's 5′-LTR instead of NIC as an alternative *ECL* terminator in Col-0 coopted an EDM2-dependent epigenetic mechanism to the control of *RPP7* expression. This epigenetic mechanism may have equipped its hosts with a selective advantage, as it provided an additional switch to fine-tune levels of RPP7-coding transcripts in response to pathogen recognition.

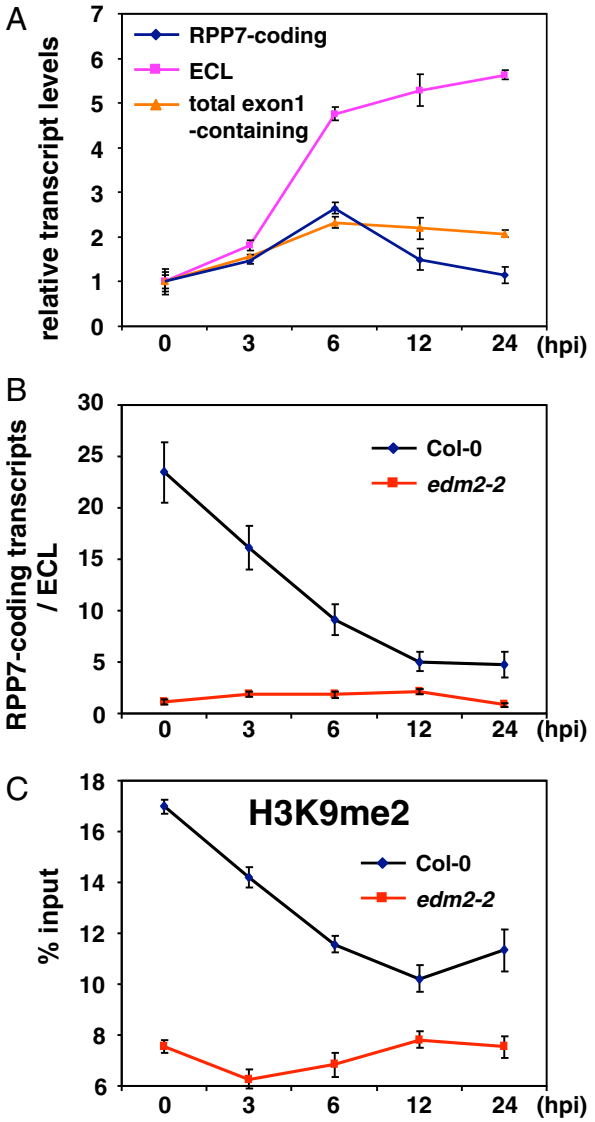

Fig. 6. Fine-tuning of RPP7-coding transcript levels in response to *Hpa*Hiks1 infection. (*A*) qRT-PCRs to monitor levels of RPP7-coding, ECL, and total exon1-containing transcripts in Col-0 plants at various time points after infection with *Hpa*Hiks1. (*B*) Ratio between RPP7-coding and ECL transcripts at the time points after infection with *Hpa*Hiks1 in Col-0 and *edm2-2* plants. (*C*) Levels of H3K9me2 at region 6 in Fig. 3*A* at the time points after infection with *Hpa*Hiks1 in Col-0 and *edm2-2* plants. Error bars represent SD or SEM for transcripts or H3K9me2, respectively, of three qPCR values from one representative of three independent experiments. Independent replications of the time course reproducibly showed the same trend of changes of transcripts and H3K9me2 levels. However, the timing of induction varied depending on the experiment.

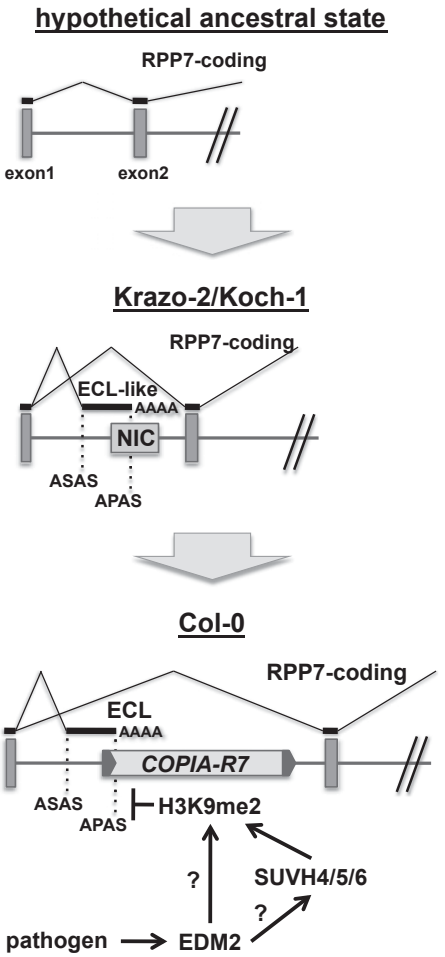

Fig. 7. Model for the cooption of H3K9me2-mediated APA regulation to the *RPP7* locus through *COPIA-R7* domestication. The starting point in the evolution of the *RPP7*/*COPIA-R7* haplotype is a hypothetical ancestral state in a putative common ancestor of Col-0, Krazo2, and Koch1. Here, *RPP7* is not subject to sophisticated co- or posttranscriptional expression control. An intermediate example is the state represented in Koch1 and Krazo2, where an alternative polyadenylation mechanism involving an alternative polyadenylation site (APAS) provided by the NIC sequence defines the balance between RPP7-coding and ECL-like transcripts. An advanced state, that likely has evolved from the intermediate state represented in Koch1 and Krazo2, is the situation in Col-0 (and likely most other *A. thaliana* accessions) where, due to the *COPIA-R7* domestication event, an EDM2-controlled and H3K9me2-dependent alternative polyadenylation mechanism allows for dynamic regulation of the RPP7-coding transcript/ECL balance in response to defense-related stimuli. In the absence of functional EDM2 or SUVH4/5/6, the chromatin region of *ECL*/*COPIA-R7* in *RPP7* is marked only by low levels of H3K9me2. Under this condition, the promoter-proximal APAS, which is present in 5′-LTR of *COPIA-R7*, is preferentially used. Enhanced use of this APAS results in enhanced levels of ECL and reduced levels of RPP7-coding transcripts. Function of EDM2 or SUVH4/5/6 counteracts this negative effect on RPP7-coding transcript levels by enhancing H3K9me2 in the *ECL*/*COPIA-R7* region. Increased H3K9me2 levels are possibly associated with changes of the local chromatin structure or altered levels of other epigenetic marks. It is unknown whether EDM2 directly affects H3K9me2 at *RPP7*/*COPIA-R7* or whether it acts via histone methyl transferases, such as SUVH4, -5, and/or -6. ASAS, alternative splicing acceptor site. Dark gray boxes represent exons. Transcript parts retained in the RPP7-coding and ECL–like transcripts are represented by black horizontal lines.

Materials and Methods

***Arabidopsis* Mutants and Natural Accessions.** *A. thaliana* accessions Col-0 (CS70000), Krazo-2 (CS76422), and Koch-1 (CS76396) were obtained from the *Arabidopsis* Biological Resource Center (ABRC, Ohio State University). The *edm2-2*, *edm2-3*, and *edm2-4* alleles have been described previously (23). The *suvh4 suvh5 suvh6* triple mutant (*suvh456*) was kindly provided by Judith Bender (Brown University, Providence, RI). All mutant plants used are exclusively in the Col-0 background.

Transgenic Lines. The transgenic complementation lines ($E2_{pro}$*:HA-E2*) were generated as described previously (24). To generate EDM2 silencing lines in *Arabidopsis* natural variants, a DNA trigger sequence was cloned into the pJawohl8-RNAi vector (GenBank AF408413; kindly provided by Imre E. Somssich, Max Planck Institute for Plant Breeding Research, Cologne, Germany), and plants were transformed with the resulting pJawohl8-EDM2 binary vector by the floral dipping method (36). Additional experimental procedures are available in *SI Materials and Methods*.

qRT-PCR Analysis. Total RNA was isolated using TRIzol reagent (Life Technologies) and reverse-transcribed with Maxima reverse transcriptase (Thermo Scientific) and oligo$(dT)_{18}$ or gene-specific primers. Real-time PCR was performed with the MyiQ detection system (Bio-Rad). The primers used for RT-PCRs are listed in Table S1. Details are described in *SI Materials and Methods*.

Chromatin Immunoprecipitation. Details of ChIP procedures are described in *SI Materials and Methods*. We used commercially available antibodies specific to dimethyl H3K9 (Wako Pure Chemical Industries; 308–32361), monomethyl and trimethyl H3K27 (Millipore; 07–448, lot no. 2019561; Abcam; ab6002), trimethyl H3K4 (Millipore; 05–1339), HA-tag (Abcam; ab9110), Histone H3 C-terminal (Active Motif; 61277), or RNAPII (Abcam; ab5408). The primers used for ChIPs are listed in Table S1.

RACE. The 5′– and 3′RACE were performed using GeneRacer kit (Life Technologies), according to the manufacturer's instruction. The primers used for RACEs are listed in Table S1.

Genomic DNA and cDNA Cloning in *Arabidopsis* Natural Accessions. The genomic DNA fragments of $RPP7_{krazo\text{-}2}$ and $RPP7_{Koch\text{-}1}$ were PCR-amplified with primers designed using the Col-0 *RPP7* as the reference. These genomic fragments were sequenced and assembled based on overlapping sequences. The regions including junctions between the PCR fragments were then PCR-amplified using genomic DNA as template to verify their continuity in the genomes. cDNAs of $RPP7_{krazo\text{-}2}$ and $RPP7_{Koch\text{-}1}$ were cloned by PCR with primers designed with the revealed genomic sequences. cDNAs of $EDM2_{krazo\text{-}2}$ and $EDM2_{Koch\text{-}1}$ were PCR-amplified with primers designed using the cDNA sequence of Col-0 *EDM2* as the reference. The primers used for genomic DNA and cDNA cloning are listed in Table S1.

Hyaloperonospora *Arabidopsidis* Infection. The *H. arabidopsidis* isolate Hiks1 (*Hpa*Hiks1) was described previously (37, 38). *Hpa*Hiks1 was grown, propagated, and applied to *Arabidopsis* plants as described previously (39). Using Preval sprayers, 1-wk-old seedlings were spray-inoculated with spore suspensions of *Hpa*Hiks1 (5×10^4 spores per mL). Plants were scored at 7 d after infection for severity of infection by lactophenol trypan blue staining (40).

ACKNOWLEDGMENTS. We thank J. N. Bailey-Serres and Mercedes Schroeder (both University of California, Riverside) for critical reading of the manuscript; I. E. Somssich (Max Planck Institute for Plant Breeding Research) for the pJawohl8-RNAi vector; J. Bender (Brown University) for providing the *suvh456* mutant; E. Holub (Warwick HRI) for providing the *H. arabidopsidis* isolates Hiks1; the *Arabidopsis* Biological Resource Center stock center for providing *Arabidopsis* natural accessions and T-DNA mutants; and the *A. thaliana* 1001 Genomes Project for genome sequencing data. This work was supported by National Science Foundation Grant IOS 1052556 (to T. E.).

1. Rigal M, Mathieu O (2011) A "mille-feuille" of silencing: Epigenetic control of transposable elements. *Biochim Biophys Acta* 1809(8):452–458.
2. Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416(6880): 556–560.
3. Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks: DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12(16):1360–1367.
4. Ebbs ML, Bartee L, Bender J (2005) H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. *Mol Cell Biol* 25(23):10507–10515.
5. Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell* 18(5):1166–1176.
6. Jacob Y, et al. (2009) ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol* 16(7):763–768.
7. Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr Opin Genet Dev* 15(6):621–627.
8. Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: Lessons from plants. *Curr Opin Genet Dev* 18(2):188–192.
9. Dooner HK, Weil CF (2007) Give-and-take: Interactions between DNA transposons and their host plant genomes. *Curr Opin Genet Dev* 17(6):486–492.
10. Lisch D (2013) How important are transposons for plant evolution? *Nat Rev Genet* 14(1):49–61.
11. Orphanides G, Reinberg D (2002) A unified theory of gene expression. *Cell* 108(4): 439–451.
12. Wu X, et al. (2011) Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. *Proc Natl Acad Sci USA* 108(30):12533–12538.
13. Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. *Mol Cell* 43(6):853–866.
14. Millevoi S, Vagner S (2010) Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. *Nucleic Acids Res* 38(9):2757–2774.
15. Pandya-Jones A, Black DL (2009) Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15(10):1896–1908.
16. de la Mata M, Lafaille C, Kornblihtt AR (2010) First come, first served revisited: Factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA* 16(5):904–912.
17. Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136(4):688–700.
18. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* 144(1):16–26.
19. Spies N, Nielsen CB, Padgett RA, Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* 36(2):245–254.
20. Mavrich TN, et al. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* 18(7):1073–1083.
21. Lian Z, et al. (2008) A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3′ end RNA polyadenylation. *Genome Res* 18(8): 1224–1237.
22. Maekawa T, Kufer TA, Schulze-Lefert P (2011) NLR functions in plant and animal immune systems: So far and yet so close. *Nat Immunol* 12(9):817–826.
23. Eulgem T, et al. (2007) EDM2 is required for RPP7-dependent disease resistance in Arabidopsis and affects RPP7 transcript levels. *Plant J* 49(5):829–839.
24. Tsuchiya T, Eulgem T (2013) Mutations in EDM2 selectively affect silencing states of transposons and induce plant developmental plasticity. *Sci Rep* 3:1701.
25. Xiao S, Brown S, Patrick E, Brearley C, Turner JG (2003) Enhanced transcription of the Arabidopsis disease resistance genes RPW8.1 and RPW8.2 via a salicylic acid-dependent amplification circuit is required for hypersensitive cell death. *Plant Cell* 15(1):33–45.
26. Yi H, Richards EJ (2007) A cluster of disease resistance genes in Arabidopsis is coordinately regulated by transcriptional activation and RNA silencing. *Plant Cell* 19(9):2929–2939.
27. Holt BF, 3rd, Belkhadir Y, Dangl JL (2005) Antagonistic control of disease resistance protein stability in the plant immune system. *Science* 309(5736):929–932.
28. Li F, et al. (2012) MicroRNA regulation of plant innate immune receptors. *Proc Natl Acad Sci USA* 109(5):1790–1795.
29. Bomblies K, et al. (2007) Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol* 5(9):e236.
30. Tsuchiya T, Eulgem T (2011) EMSY-like genes are required for full RPP7-mediated race-specific immunity and basal defense in Arabidopsis. *Mol Plant Microbe Interact* 24(12):1573–1581.
31. Liu F, Marquardt S, Lister C, Swiezewski S, Dean C (2010) Targeted 3′ processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science* 327(5961): 94–97.
32. Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annu Rev Genet* 33:479–532.
33. Cao J, et al. (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nat Genet* 43(10):956–963.
34. Tsuchiya T, Eulgem T (2010) Co-option of EDM2 to distinct regulatory modules in Arabidopsis thaliana development. *BMC Plant Biol* 10:203.
35. Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res* 12(7): 1075–1079.
36. Clough SJ, Bent AF (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16(6):735–743.
37. Holub EB, Beynon JL, Crute IR (1994) Phenotypic and genotypic characterization of interactions between isolates of Peronospora parasitica and accessions of Arabidopsis thaliana. *MPMI-Molecular Plant Microbe Interactions* 7(2):223–239.
38. Parker JE, et al. (1993) Phenotypic characterization and molecular mapping of the Arabidopsis thaliana locus RPP5, determining disease resistance to Peronospora parasitica. *Plant J* 4(5):821–831.
39. McDowell JM, et al. (2000) Downy mildew (Peronospora parasitica) resistance genes in Arabidopsis vary in functional requirements for NDR1, EDS1, NPR1 and salicylic acid accumulation. *Plant J* 22(6):523–529.
40. Koch E, Slusarenko A (1990) Arabidopsis is susceptible to infection by a downy mildew fungus. *Plant Cell* 2(5):437–445.
41. Tsuchiya T, Eulgem T (2010) The Arabidopsis defense component EDM2 affects the floral transition in an FLC-dependent manner. *Plant J* 62(3):518–528.