A domesticated Harbinger transposase forms a complex with HDA6 and promotes histone H3 deacetylation at genes but not TEs in Arabidopsis

JIPB Journal of Integrative Plant Biology

Research Article
https://doi.org/10.1111/jipb.13108

A domesticated *Harbinger* transposase forms a complex with HDA6 and promotes histone H3 deacetylation at genes but not TEs in *Arabidopsis*[oo]

Xishi Zhou[1†], Junna He[2,3†], Christos N. Velanis[4†], Yiwang Zhu[1], Yuhan He[1], Kai Tang[3,5], Mingku Zhu[3,6], Lisa Graser[4,7], Erica deLeau[4], Xingang Wang[3], Lingrui Zhang[3], W. Andy Tao[8], Justin Goodrich[4*], Jian-Kang Zhu[5*] and Cui-Jun Zhang[1,5*]

1. Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China
2. College of Horticulture, China Agricultural University, Beijing 100193, China
3. Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907, USA
4. Institute of Molecular Plant Science, School of Biological Sciences, University of Edinburgh, Daniel Rutherford Building, Max Born Crescent, Edinburgh EH9 3BF, United Kingdom
5. Shanghai Center for Plant Stress Biology and CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 201602, China
6. Institute of Integrative Plant Biology, School of Life Sciences, Jiangsu Normal University, Xuzhou 221116, China
7. University of Applied Sciences Mannheim, Paul-Wittsack-Str. 10, Mannheim 68163, Germany
8. Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA

[†]These authors contributed equally to this work.

*Correspondences: Justin Goodrich (Justin.Goodrich@ed.ac.uk); Jian-Kang Zhu (jkzhu@psc.ac.cn); Cui-Jun Zhang (zhangcuijun@caas.cn, Dr. Zhang is responsible for the distribution of the materials associated with this article)

Xishi Zhou

Cui-Jun Zhang

ABSTRACT

In eukaryotes, histone acetylation is a major modification on histone N-terminal tails that is tightly connected to transcriptional activation. HDA6 is a histone deacetylase involved in the transcriptional regulation of genes and transposable elements (TEs) in *Arabidopsis thaliana*. HDA6 has been shown to participate in several complexes in plants, including a conserved SIN3 complex. Here, we uncover a novel protein complex containing HDA6, several *Harbinger* transposon-derived proteins (HHP1, SANT1, SANT2, SANT3, and SANT4), and MBD domain-containing proteins (MBD1, MBD2, and MBD4). We show that mutations of all four *SANT* genes in the *sant-null* mutant cause increased expression of the flowering repressors *FLC*, *MAF4*, and *MAF5*, resulting in a late flowering phenotype. Transcriptome deep sequencing reveals that while the SANT proteins and HDA6 regulate the expression of largely overlapping sets of genes, TE silencing is unaffected in *sant-null* mutants. Our global histone H3 acetylation profiling shows that SANT proteins and HDA6 modulate gene expression through deacetylation. Collectively, our findings suggest that *Harbinger* transposon-derived SANT domain-containing proteins are required for histone deacetylation and flowering time control in plants.

Keywords: deacetylation, flowering, Harbinger, HDA6, histone acetylation, protein complex, SANT

Zhou, X., He, J., Velanis, C.N., Zhu, Y., He, Y., Tang, K., Zhu, M., Graser, L., de Leau, E., Wang, X., Zhang, L., Andy Tao, W., Goodrich, J., Zhu, J.-K., and Zhang, C.J. (2021) A domesticated *Harbinger* transposase forms a complex with HDA6 and promotes histone H3 deacetylation at genes but not TEs in *Arabidopsis*. J. Integr. Plant Biol. **00:** 1–13.

INTRODUCTION

In eukaryotic cells, histones and DNA are packed and ordered into highly structured units called nucleosomes. Nucleosomal histones are subject to multiple posttranslational modifications, such as methylation, acetylation, phosphorylation, and ubiquitination, which play critical roles in various biological processes that include transcriptional regulation, chromosome packaging, and repair of DNA damage (Bannister and Kouzarides, 2011; Zhao et al., 2019; Shim et al., 2020). Acetylation of histones occurs at specific lysine residues on their N-terminal tails and is associated with transcriptional activation (Struhl, 1998). Histone acetyltransferases (HATs) and histone deacetylases (HDAs or HDACs) regulate histone acetylation status by adding or removing acetylation marks, respectively (Marmorstein and Zhou, 2014; Seto and Yoshida, 2014; Guo et al., 2021).

The *Arabidopsis* (*Arabidopsis thaliana*) genome encodes 18 annotated HDACs that fall into three families based on sequence: the homologs of yeast (*Saccharomyces cerevisiae*) REDUCED POTASSIUM DEPENDENCY 3 (RPD3), the plant-specific HISTONE DEACETYLASE 2 (HD2), and the Sir2-like family (Hollender and Liu, 2008; Chen et al., 2020; Yruela et al., 2021). HDA6, an RPD3-type HDAC, has been extensively studied in *Arabidopsis*. HDA6 is involved in various biological processes such as transcriptional gene silencing (Probst et al., 2004; Earley et al., 2010; To et al., 2011; Liu et al., 2012; Yu et al., 2017; Yang et al., 2020), responses to drought and salt stress (Chen et al., 2010; Kim et al., 2017), expression of circadian clock genes (Hung et al., 2018), mRNA polyadenylation (Lin et al., 2020), and flowering (Wu et al., 2008; Yu et al., 2011; Ning et al., 2019). The *hda6* mutant shows late flowering phenotype due to the upregulation and hyperacetylation of the flowering repressor loci *FLOWERING LOCUS C* (*FLC*), *MADS AFFECTING FLOWERING 4* (*MAF4*), and *MAF5* (Wu et al., 2008; Yu et al., 2011; Ning et al., 2019). In *Arabidopsis*, FVE and FLOWERING LOCUS D (FLD) interact with HDA6 to repress *FLC*, *MAF4* and *MAF5* expression through histone deacetylation and demethylation (Gu et al., 2011; Yu et al., 2011). Recently, HISTONE DEACETYLATION COMPLEX 1 (HDC1), SIN3-LIKE proteins (SNLs), and MULTICOPY SUPRESSOR OF IRA 1 (MSI1) were reported to be shared subunits of the HDA6 and HDA19 complexes (Ning et al., 2019). Compared to *hda6* mutants, which flower late in both long days (LD) and short days (SD), the *hda19*, *hdc1*, *snl2 snl3 snl4*, and *msi1* mutants show delayed flowering time in LD only (Steinbach and Hennig, 2014; Ning et al., 2019), indicating that HDC1, SNLs, and MSI1 may not participate in the HDA6-mediated control of flowering. Notably, transcriptome analysis indicated that HDA6 is involved in the transcriptional regulation of many genes and transposable elements (TEs) (Yu et al., 2016). It was later reported that HDA6 interacts with the histone H3 lysine 9 methyltransferases SU(VAR)3-9 HOMOLOG 4 (SUVH4), SUVH5, and SUVH6 to co-regulate the silencing of TEs (Yu et al., 2017). However, relatively little is known about how HDA6 specifically regulates gene transcription.

Transposable elements are repetitive sequences that constitute a large part of plant and animal genomes. Although usually considered as selfish or parasitic, TEs are important sources of emerging new genes (Sinzelle et al., 2009; Bourque et al., 2018). *PIF*/*Harbinger* class transposons usually encode two proteins: a nuclease and a SANT/myb/trihelix domain-containing DNA binding protein (Kapitonov and Jurka, 2004; Zhang et al., 2004). Interaction between these two proteins is required to reconstitute transposase activity and effective transposition of *PIF*/*Harbinger* elements (Sinzelle et al., 2008; Hancock et al., 2010). Several genes that originated by domestication of *PIF*/*Harbinger* transposases have been reported to play critical roles in regulating gene expression (Liang et al., 2015; Duan et al., 2017; Velanis et al., 2020). For instance, *Arabidopsis* HARBINGER DERIVED PROTEIN 1 (HDP1) and HDP2, a pair of anti-silencing factors derived from *Harbinger* transposons, can interact with one another and also associate with INCREASED DNA METHYLATION 1 (IDM1), IDM2, IDM3 as well as METHYL-CPG-BINDING DOMAIN 7 (MBD7) to form a histone acetyltransferase complex that prevents hypermethylation and silencing of specific loci (Duan et al., 2017). ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1 (ALP1) and ALP2 are another pair of domesticated *Harbinger*-derived proteins that antagonize silencing mediated by Polycomb group (PcG) proteins (Liang et al., 2015; Velanis et al., 2020). ALP1 and ALP2 associate with MSI1, a core component of the H3K27me3 histone methyltransferase complex Polycomb repressive complex 2 (PRC2), and may displace the accessory components EMBRYONIC FLOWER 1 (EMF1) and LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) to form a variant PRC2 complex (Velanis et al., 2020). Phylogenetic analysis indicates that there are at least four other domesticated *Harbinger*-related nucleases in *Arabidopsis* (Duan et al., 2017). However, their partners are unknown and it is unclear whether they are also part of histone modifying complexes.

In this study, by using immunoprecipitation combined with mass spectrometry (IP-MS), we found that HDA6 associates with *Harbinger* transposon-derived proteins (HHP1, SANT1, SANT2, SANT3 and SANT4), and MBD proteins (MBD1, MBD2, and MBD4) in *Arabidopsis*. Split luciferase assays in *Nicotiana benthamiana* leaves showed that HHP1 and MBD interact directly with HDA6 and SANT proteins. However, HDA6 and SANT proteins did not interact with one another directly, suggesting that their association is mediated by their interactions with HHP1 and MBD proteins. Mutations of all four *Arabidopsis SANT* genes increased the expression of the key floral repressors *FLC*, *MAF4*, and *MAF5*, resulting in a late flowering phenotype similar to that of *hda6* mutants. However, the *hhp1* single mutant and *mbd1/2/4* triple mutant had normal flowering times. Transcriptome analysis in the wild type and various mutant backgrounds revealed that HDA6, HHP1, MBD, and SANT proteins co-regulate the expression of largely overlapping groups of

genes involved in biotic and abiotic stress responses, with inactivation of HDA6 or SANT proteins having the strongest phenotypes at the molecular level. Furthermore, genome-wide histone H3 acetylation profiling revealed that SANT1, SANT2, SANT3, and SANT4 play important parts in mediating deacetylation of histone H3. Taken together, our data uncover a critical role for *Harbinger* transposon-derived SANT domain-containing proteins in histone deacetylation and flowering time control in plants.

RESULTS

HHP1 has a function distinct from that of its paralog *ALP1*

We previously reported that the *Arabidopsis* genome contains one paralog for *ALP1*, *At3g55350* (Liang et al., 2015), which we refer to hereafter as *HDA6-associated Harbinger transposon-derived protein 1* (*HHP1*) (see below). We obtained a mutant allele, *hhp1-1* (Salk_122829), with a T-DNA insertion in the second exon predicted to introduce a premature stop codon in the *HHP1* coding sequence. The *hhp1-1* mutant did not show any obvious morphological abnormalities when grown under standard conditions (Figure S1A). Given the sequence similarity between *ALP1* and *HHP1*, we postulated that the two genes might function redundantly. However, the *hhp1-1 alp1-1* double mutant exhibited the same phenotype as the *alp1-1* single mutant (Liang et al., 2015; Figure S1A). Mutation of *alp1* partially suppressed the early flowering and leaf curling phenotypes observed in the PcG mutant *lhp1*, whereas *hhp1-1 lhp1-3* double mutant plants were phenotypically similar to the *lhp1-3* single mutant (Figure S1B, C). In addition, the morphology and flowering time of the *lhp1-3 alp1-1 hhp1-1* triple mutant were comparable to those of the *lhp1-3 alp1-1* double mutant (Figure S1B, C). Together, these results indicated that HHP1 is unlikely to function like ALP1 as a component of the PRC2 complex.

Harbinger transposon-derived proteins co-purify with HDA6 and MBD proteins

HDP1 and ALP1 are *Harbinger* transposase-derived proteins and function as components of two distinct chromatin modifying complexes (Liang et al., 2015; Duan et al., 2017; Velanis et al., 2020). Since HHP1 is also a *Harbinger* transposon-derived protein (Liang et al., 2015), we next undertook a proteomics approach to test whether HHP1, like ALP1 and HDP1, is part of a chromatin modifying complex. To determine the function of HHP1, we performed immunoprecipitation followed by mass spectrometry (IP-MS) with transgenic plants overexpressing a translational fusion between HHP1 and the green fluorescent protein (*HHP1-GFP*). We identified HDA6, three MBD proteins (MBD1, MBD2, and MBD4), and four closely related SANT domain-containing proteins among the HHP1-GFP immunoprecipitated products in samples from the transgenic plants, but not the controls (Table 1). Based on its association with HDA6, we designated the ALP1 paralog HDA6-associated *Harbinger* transposon-derived protein 1 (HHP1); we named the four SANT domain-containing proteins SANT1, SANT2, SANT3, and SANT4 (Table 1). To confirm the association of these proteins, we performed a reciprocal IP-MS experiment using *HDA6-Flag* transgenic plants. In agreement with the initial IP-MS results, HHP1, SANT proteins (SANT1, SANT2, SANT3, and SANT4) and MBD proteins (MBD1 and MBD4) co-purified with HDA6, confirming that HDA6 and HHP1 interact with the same set of proteins (Table 1).

As an independent test of interaction between these proteins, we performed split luciferase assays in *N. benthamiana* leaves. We established that HHP1 and MBD proteins (MBD1, MBD2, and MBD4) form homodimers and heterodimers (Figure 1A, B and Figure S2A, B). Similarly, HHP1 and MBD proteins (MBD1, MBD2, and MBD4) interacted with HDA6 and SANT3 and more weakly with SANT1, SANT2, and SANT4 (Figure 1A, B and Figure S2A, B). No interactions were detected between HDA6 and SANT proteins (SANT1, SANT2, SANT3, and SANT4) (Figure 1 and Figure S2). These

Table 1. List of proteins co-immunoprecipitating with HHP1 and HDA6. The interacting proteins were affinity-purified and analyzed by mass spectrometry

		HHP1-IP (Rep. 1)		HHP1-IP (Rep. 2)		HHP1-IP (Rep. 3)		HDA6-IP	
Accession	Protein	Coverage (%)	Unique peptides	Coverage (%)	Unique peptides	Coverage (%)	Unique peptides	Coverage (%)	Unique peptides
AT5G63110	HDA6	62.63	18	53.29	17	68.15	26	88.96	38
AT3G55350	HHP1	74.63	27	74.88	27	89.41	35	33.74	10
AT2G47820	SANT3	39.01	27	36.52	26	52.8	37	24.10	16
AT1G55050	SANT4	23.06	19	32.79	23	37.16	31	9.62	9
AT1G09050	SANT1	28.71	9	42.14	17	45.52	17	17.25	5
AT1G09040	SANT2	23.27	7	37.43	12	45.23	19	14.05	3
AT3G63030	MBD4	51.08	8	30.11	5	76.88	11	19.35	3
AT4G22745	MBD1	47.55	8	19.12	3	56.37	11	21.57	4
AT5G35330	MBD2	37.50	6	50.00	7	66.18	10	0.00	0

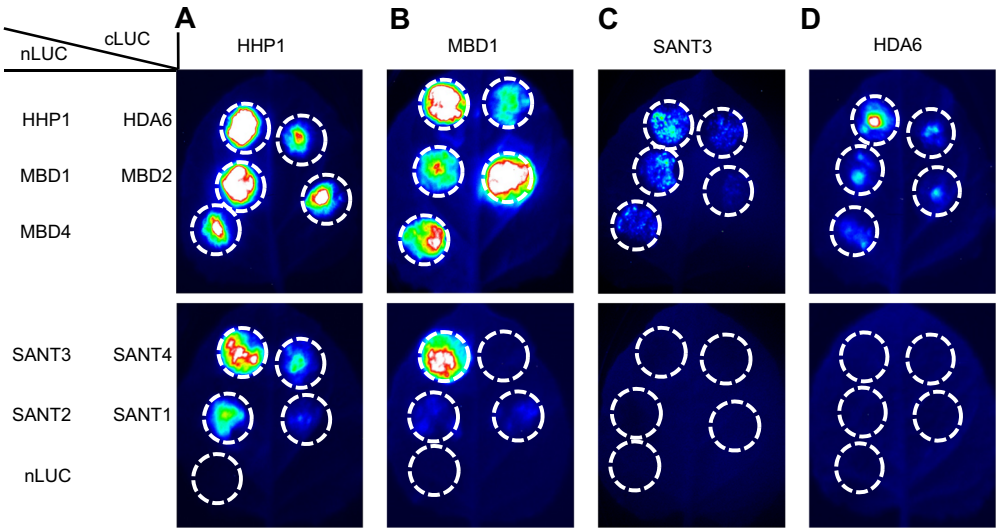

Figure 1. Assessment of protein interaction potential between HHP1, HDA6, MBD, MBD2, MBD4 and SANT1, SANT2, SANT3, and SANT4 proteins by split luciferase assays

(A) Luciferase complementation imaging (LCI) assay showing that HHP1 interacts strongly with itself, MBD1, MBD2, MBD4, and SANT3 and weakly with HDA6 and SANT2/4. **(B)** LCI assay showing that MBD1 interacts strongly with HHP1, MBD1, MBD2, MBD4, and SANT3 and weakly with HDA6. **(C)** LCI assay showing that SANT3 interacts weakly with HHP1 and MBD1. **(D)** LCI assay showing that HDA6 interacts weakly with HHP1, HDA6, MBD1, MBD2, and MBD4, but not with SANT1–SANT4. nLUC represents the empty nLUC vector.

results suggested that HHP1 and/or the MBD proteins may act as scaffolds for the association with the HDA6 and SANT proteins in one or more complexes.

In both published examples of domesticated *Harbinger* transposases in *Arabidopsis*, the nuclease-derived component (HDP1 or ALP1) retained its cooperation with the protein derived from the DNA-binding domain (HDP2 or ALP2) (Duan et al., 2017; Velanis et al., 2020). The ancestral *Harbinger* DNA-binding component shows limited similarity to SANT/Myb/trihelix motif proteins (Kapitonov and Jurka, 2004; Duan et al., 2017). As the SANT proteins identified here not only bear a SANT motif (Figure S3A) but also, in most cases (SANT2, SANT3, SANT4), can interact directly with the domesticated *Harbinger*-related nuclease HHP1, we hypothesized that HHP1 and SANTs arose via the co-domestication of ancestral *Harbinger* transposase nuclease and DNA binding components. In agreement with this hypothesis, protein sequence alignments revealed conservation across the SANT/Myb/trihelix domains of the four SANT proteins and the DNA-binding component of *Harbinger* transposases (Figure S4). In addition, HHP1 lacked the DDE motif required for nuclease activity (Liang et al., 2015), while the SANT proteins lacked two of the three conserved tryptophan residues necessary for DNA binding activity (Figure S4), consistent with domestication and acquisition of novel functions.

SANT1, SANT2, SANT3, and SANT4 promote flowering in *Arabidopsis*

Because SANT1/2/3/4, HHP1, and MBD1/2/4 physically associate with HDA6, we asked whether any of these proteins share common functions with HDA6 in *Arabidopsis*. BLAST analysis revealed that SANT1, SANT2, SANT3, and SANT4 are more closely related to each other than to any other *Arabidopsis* protein with a SANT domain (Figure S3B), indicating that they may function redundantly. We first obtained a mutant allele (Salk_031353) harbouring a T-DNA insertion in the promoter of *SANT3* (Figure S5A). To explore redundancy between SANT family members, we generated higher-order *sant* mutants in the Salk_031353 background using a multiplex CRISPR/Cas9 gene editing system (Zhang et al., 2016) with guide RNAs that targeted all four genes (Figure S5A). Through genotyping, we first identified a higher order mutant bearing large deletions in *SANT1*, *SANT2*, and *SANT4* (Figure S5B). However, subsequent RT-PCR analysis showed that the expression of *SANT3* was reduced but not eliminated by the Salk_031353 T-DNA insertion and we therefore termed this quadruple mutant *sant-weak* (Figure S5C). After screening further we identified *sant-null*, containing an additional frameshift mutation causing early translation termination in *SANT3*, as well as the large deletions in *SANT1*, *SANT2*, and *SANT4* (Figure S5B, D). When grown under long day conditions, the *sant-weak* mutant flowered slightly later than Col-0 (Figure S6), while the *sant-null* mutant showed an obviously late flowering phenotype, but less severe than that of the *hda6* mutant (Figure 2A, B). To test if HHP1 and MBDs are also required for normal flowering in *Arabidopsis*, we obtained a likely null *hhp1* mutant allele (*hhp1-2*) via CRISPR/Cas9 genome editing and a *mbd1/2/4* triple mutant (Figure S7). However, both the *hhp1-2* single mutant and the *mbd1/2/4* triple mutant flowered at the same time as Col-0 (Figure 2A, B), arguing against the involvement of these genes in flowering time control in *Arabidopsis*.

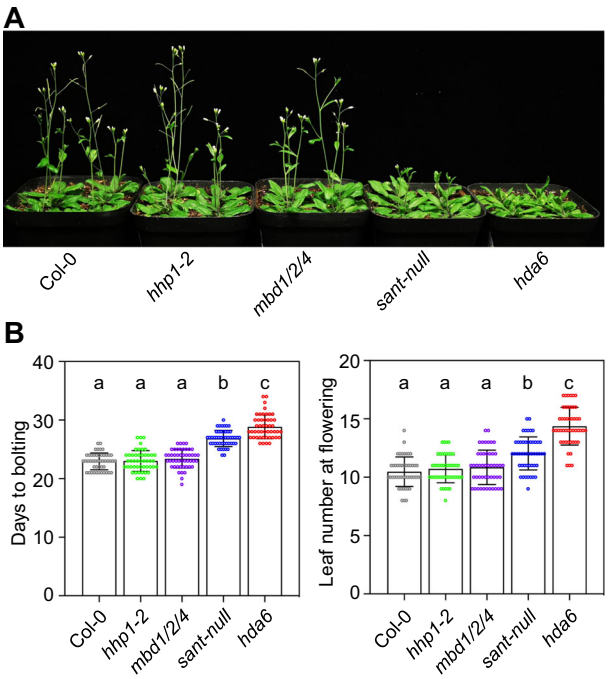

Figure 2. Flowering phenotypes of Col-0, *hhp1*, *mbd1/2/4*, *sant-null*, and *hda6* plants under long-day conditions

(A) The *sant-null* and *hda6* mutants flower late when grown in long-day conditions, while the *hhp1-2* and *mbd1/2/4* mutants flower at the same time as Col-0. **(B)** Days to bolting (left) and number of rosette leaves upon bolting (right) for Col-0, *hhp1-2*, *mbd1/2/4*, *sant-null*, and *hda6* mutants grown in long-day conditions. At least 40 plants were scored for each genotype. The *y*-axes denote days to bolting (left) and rosette leaf number (right). Each dot is a separate plant. Different letters above bars indicate statistically significant differences (ANOVA; Duncan's multiple range test, $P < 0.05$).

SANT1/2/3/4 and HDA6 co-regulate the expression of flowering repressors

Several studies have shown that the late flowering phenotype of *hda6* is the result of increased expression of the flowering repressors *FLC*, *MAF4*, and *MAF5* (Wu et al., 2008; Yu et al., 2011; Ning et al., 2019). To further study the function of SANT1, SANT2, SANT3, SANT4, HHP1, MBD1, MBD2, and MBD4 in transcriptional regulation of genes related to flowering, we performed transcriptome deep sequencing (RNA-seq) of 12-day-old Col-0, *sant-null*, *hhp1-2*, *mbd1/2/4*, and *hda6* seedlings grown in long days. A visual inspection of RNA-seq reads mapping to each locus in the Integrated Genome Viewer (IGV) browser showed that *FLC*, *MAF4*, and *MAF5* are upregulated strongly in the *sant-null* and *hda6* mutants, and to a lesser extent in *hhp1-2* and *mbd1/2/4*, relative to Col-0 seedlings (Figure 3A–C). RT-qPCR analyses confirmed that all three genes are upregulated in the *sant-null* and *hda6* mutants, and much more modestly so in *hhp1-2* and *mbd1/2/4* (Figure 3D). These results indicated that SANT1/2/3/4 and HDA6 contribute to the regulation of flowering time in long days and suggested that they may act together to promote flowering.

SANT1/2/3/4 and HDA6 co-regulate global gene expression

HDA6 is a histone deacetylase that functions in transcriptional repression. We therefore tested whether SANT1/2/3/4, HHP1, and MBD1/2/4 shared functions with HDA6 in transcriptional regulation of the *Arabidopsis* genome. We identified 924 and 1,047 genes that were upregulated in the *hda6* and *sant-null* mutants, respectively, relative to Col-0 samples,

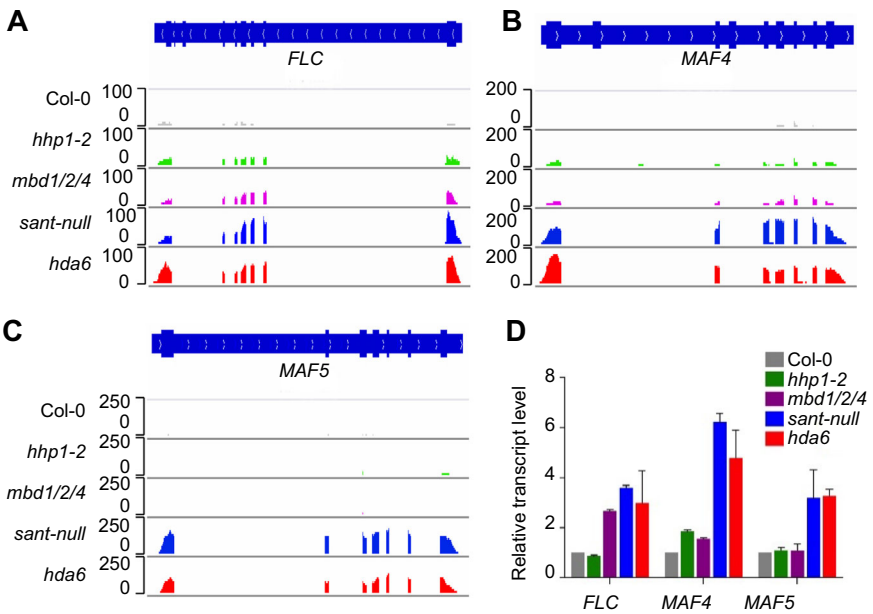

Figure 3. HDA6 and SANT1/2/3/4 co-regulate the expression of genes controlling flowering time

(A–C) Snapshots of the genome browser illustrating expression levels of *FLC*, *MAF4*, and *MAF5* in the indicated genotypes. **(D)** Relative transcript levels of *FLC*, *MAF4*, and *MAF5*, as determined by RT-qPCR, in Col-0 and *hhp1-2*, *mbd1/2/4*, *sant-null*, and *hda6* mutants. Error bars indicate the standard deviation of three biological replicates.

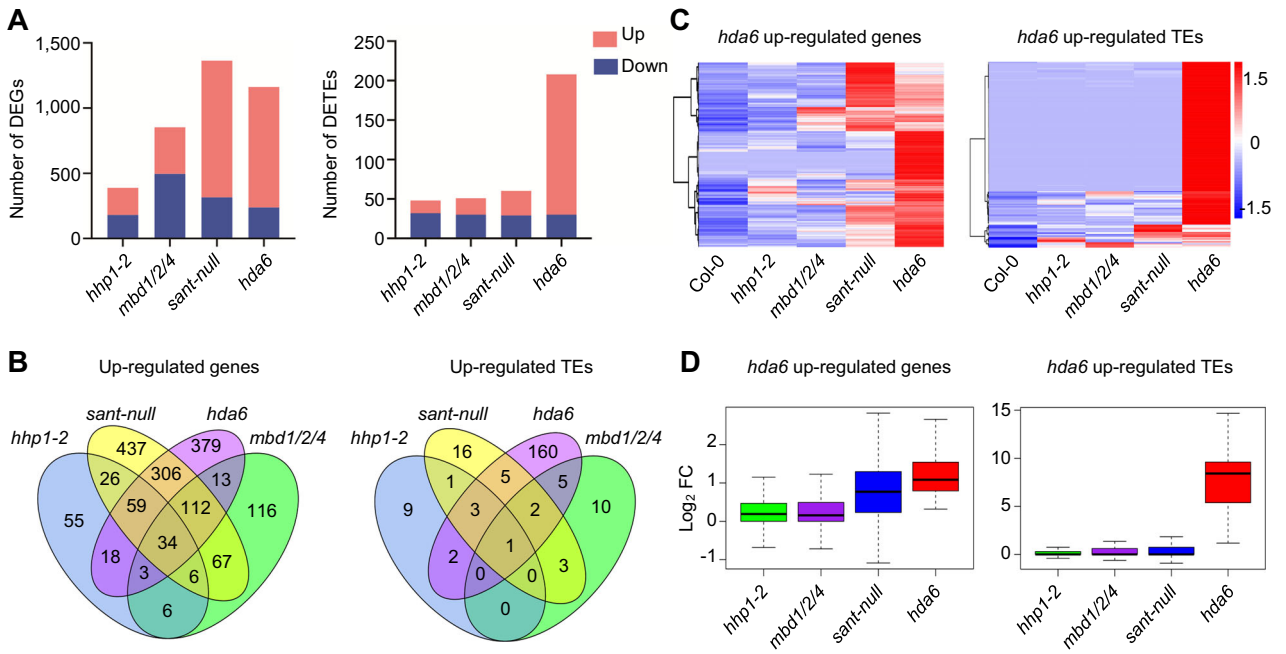

Figure 4. SANT1/2/3/4 and HDA6 co-regulate the expression of genes but not TEs

(A) Number of differentially expressed genes (left) and TEs (right) in *hhp1-2*, *mbd1/2/4*, *sant-null*, and *hda6* mutants relative to Col-0 ($P < 0.05$ and a two-fold cutoff were used). **(B)** Venn diagrams showing the overlap of upregulated genes (left) and TEs (right) in *hhp1-2*, *mbd1/2/4*, *sant-null*, and *hda6* mutants.**(C)** Heatmap illustrating the transcript levels of *hda6*-mediated upregulated genes (left) and TEs (right) in *hda6* and other indicated mutants. The color scale indicates normalized FPKM values. **(D)** Boxplot of transcript levels for *hda6*-mediated upregulated genes (left) and TEs (right) in *hda6* and indicated mutants. Each box represents the $\log_2$ (FPKM + 1 values of mutants/FPKM + 1 values of Col-0).

and another 239 and 317 that were downregulated, consistent with roles of HDA6 and the SANT proteins in repressing gene expression (Figure 4A; Dataset S1). However, we detected fewer (207 and 357) upregulated genes in the *hhp1-2* and *mbd1/2/4* mutants, respectively, relative to Col-0 (Figure 4A; Dataset S1). The relatively modest transcriptional changes in these two mutants were consistent with their weaker effects on *FLC*, *MAF4* and *MAF5* expression and flowering time relative to the *hda6* and *sant-null* mutants.

Of the 924 upregulated genes in *hda6*, 511 (55%) genes were also upregulated in the *sant-null* mutant (Figure 4B; Figure S8). Heatmap and box plot analysis showed that most of the up-regulated genes in *hda6* also appeared up-regulated in the *sant-null* mutant (Figure 4C, D), suggesting that SANT1/2/3/4 and HDA6 have similar functions. Consistently, we also found that most of the up-regulated genes in the *sant-null* mutant appeared upregulated in *hda6* (Figure S9). Although relatively fewer genes were upregulated in *hhp1-2* or in *mbd1/2/4* than in *hda6*, a high proportion (about 50%) of them were also upregulated in *hda6*; the extent of this overlap was statistically significant, as determined by a hypergeometric distribution (Figure S8). Gene ontology (GO) term enrichment analyses revealed that the genes up-regulated in *sant-null* and *hda6* mutants are involved in stress responses (Figure S10). HDA6 has been reported to be required for transcriptional silencing of TEs (Yu et al., 2017). Indeed, 178 TEs were upregulated in the *hda6* mutant relative to Col-0 in our RNA-seq data (Figure 4A; Data S2), whereas only 31, 16, and 21 TEs were upregulated in the *sant-null*, *hhp1-2*, and *mbd1/2/4* mutants, respectively (Figure 4A; Dataset S2). The expression of most of the TEs upregulated in *hda6* was not affected in the *sant-null*, *hhp1-2*, and *mbd1/2/4* mutant backgrounds (Figure 4B–D). These results suggested that even though SANT1/2/3/4, HHP1, and MBD1/2/4 form a complex with HDA6, they are not involved in maintaining TE silencing at heterochromatin. Collectively, these results indicated that SANT1/2/3/4 and HDA6 co-regulate the expression of genes, but not that of TEs.

SANT proteins are required for histone H3 deacetylation

Previous studies revealed that HDA6 is an RPD3-type HDAC critical for histone H3 deacetylation of target sites. The association of SANT proteins with HDA6 prompted us to examine their contribution to histone H3 deacetylation in *vivo*. We investigated the global landscape of histone H3 acetylation in wild-type, *sant-null* and *hda6* mutant seedlings by chromatin immunoprecipitation followed by sequencing (ChIP-seq). Based on a principal component analysis of three replicates per genotype, the results of the ChIP-seq experiment were consistent and reproducible (Figure 5A). We identified 784 regions with higher levels of histone H3 acetylation in the *hda6* mutant (Dataset S3). Compared to wild type, 2,751 regions showed significantly higher levels of histone H3 acetylation in the *sant-null* mutant (Dataset S3). It is surprising that more regions are affected in *sant-null* than in

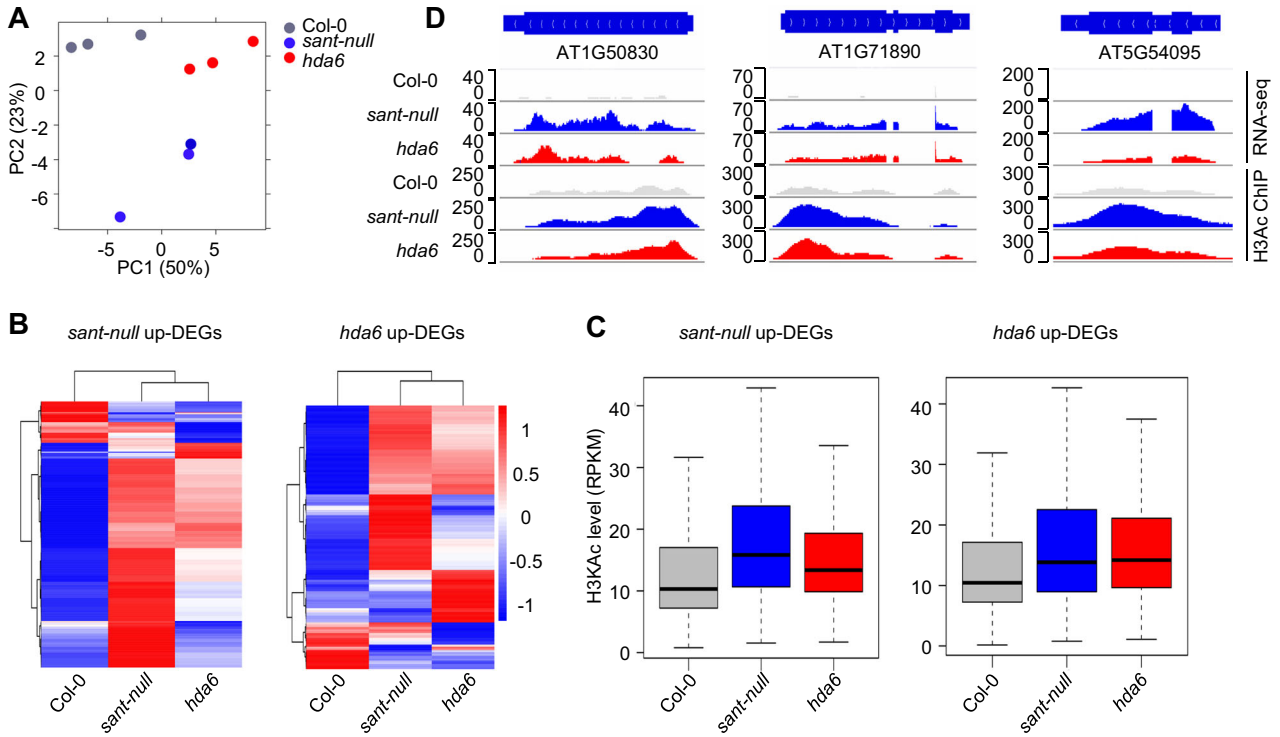

Figure 5. SANT1/2/3/4 and HDA6 co-regulate histone acetylation

(A) Principal component analysis of H3KAc-ChIP-seq data. Each dot represents one replicate of the indicated ChIP-seq sample. **(B)** Heatmap representation of H3KAc levels for upregulated genes identified by RNA-seq in *sant-null* (left) and *hda6* (right). The color scale indicates normalized RPKM values. DEG, differentially expressed gene. **(C)** Boxplot of H3KAc levels for upregulated genes identified by RNA-seq in *sant-null* (left) and *hda6* (right). Each box represents the RPKM values (reads per kilobase per million mapped reads). **(D)** Snapshots of the genome browser illustrating expression and H3KAc levels at three representative loci in Col-0, *sant-null*, and *hda6*.

hda6, especially given that HDA6 regulates TE expression and SANT1-4 do not. However, there are three genes (*HDA7, HDA9, HDA19*) that are closely related to *HDA6* in *Arabidopsis* (Pandey et al., 2002) and *HDA19* has been shown to act partially redundantly with *HDA6* in some cases (Tanaka et al., 2008). We speculated that in the absence of HDA6, HDA19 may participate in the HHP1/SANT/MBD complexes, which would result in *hda6* mutants being less severe than *sant-null* mutants. Heatmap and box plot analysis showed that the H3KAc level of *sant-null*-mediated up-regulated peaks was increased in *hda6*, but to a lesser extent (Figure S11). Our RNA-seq and RT-qPCR analyses showed that the flowering repressors *FLC*, *MAF4*, and *MAF5* were upregulated in the *sant-null* and *hda6* mutants (Figure 3). The ChIP-seq data indicated that the H3 acetylation levels of *FLC*, *MAF4*, and *MAF5* were increased in *hda6* and *sant-null* mutants (Figure S12). We further determined whether the up-regulated genes in *hda6* and *sant-null* mutant identified by RNA-seq showed increased H3KAc levels. Heatmaps and boxplots indicated that H3 acetylation levels of most genes that were upregulated in *hda6* and *sant-null* mutants were increased in *hda6* and *sant-null* mutants (Figure 5B and C; Dataset S4). However, we also found that H3 acetylation levels of a small portion of genes that were upregulated in *hda6* and *sant-null* mutants were not increased in *hda6* and *sant-null* mutants (Figure 5B; Dataset S4). This suggests that the involvement of HDA6 and SANT1/2/3/4 in transcriptional repression at some loci is independent of histone H3 deacetylation. A visual inspection of sequencing reads mapping to each locus in the IGV browser indicated that RNA-seq reads and histone H3 acetylation levels at three representative loci were significantly increased in *hda6* and *sant-null* mutants (Figure 5D). The higher transcript levels of these three genes returned to wild-type levels when the *sant-null* mutant was transformed with a genomic fragment encompassing *SANT1*, *SANT3*, or *SANT4* (Figure S5E). Together, these results demonstrated that SANT proteins play important roles in mediating deacetylation of histone H3.

DISCUSSION

Transposable elements are becoming increasingly recognized as major contributors to genetic diversity and adaptive evolutionary innovations (Oliver et al., 2013; Hosaka and Kakutani, 2018), either as a direct consequence of insertional transposition (Henaff et al., 2014; Makarevitch et al., 2015) or through the neofunctionalization of TE-encoded transposases as host genes following their molecular domestication (Bundock and Hooykaas, 2005; Jangam et al.,

2017; Cosby et al., 2021). Our study adds to the growing body of evidence supporting the recruitment of domesticated transposases by the host core epigenetic machinery as a recurrent theme in the regulation of developmental and/or environmental responses. We show that five co-domesticated proteins derived from the *Harbinger* transposon, HHP1 and SANT1–SANT4, form one or more novel histone deacetylase complexes with HDA6. Moreover, we demonstrate that these complexes are important for regulating the expression of common target genes via histone deacetylation.

HHP1 is in histone deacetylase complexes with HDA6, MBD and SANT proteins

Our reciprocal IP-MS experiments indicate that HHP1 and HDA6 interact with one another and associate with the MBD1, MBD2, MBD4, and SANT1–SANT4 proteins. These results are also supported by split luciferase assays that show numerous direct interactions between these proteins, including HHP1 with HDA6, as well as MBD1, MBD2, and MBD4 and SANT2, SANT3, and SANT4. However, HDA6 did not interact directly with the SANT proteins, suggesting that the MBD and HHP1 proteins mediate their interaction. Until the biochemical purification of the entire HDA6-containing complex(es), it is unclear whether the four SANT proteins and the three MBD proteins form one or more complexes with HHP1 and HDA6. Given that SANT1–SANT4 are closely related, as are MBD1, MBD2, and MBD4 (Zemach and Grafi, 2003), redundancy between individual SANT or MBD protein is likely, raising the possibility that various HDA6/HHP1/MBD/SANT subcomplexes may exist within cells but differ in their constituent SANT or MBD member. If these interactions are biologically relevant, a prediction is that inactivation of the different components of the complex(es), as with the *hda6* mutant or higher-order mutants in the case of *MBD1/2/4* or *SANT1/2/34* genes, will result in overlapping phenotypes. Our transcriptome analysis of 12-day-old wild-type and mutant seedlings strongly supported this hypothesis. In particular, the overlap between genes misregulated in the mutants affecting the different components was statistically significant and much higher than would be expected by chance (Figure S8). However, there are differences in phenotypic severity, with *sant-null* and *hda6* having a much greater effect on the transcriptome than *hhp1* or *mbd1/2/4* (Figure 4). The HHP1 and MBD proteins are presumably less critical for the function of the complex(es) than HDA6 and SANT1–SANT4. The split luciferase assays indicate that HHP1 and MBD proteins can each interact with both HDA6 and SANT proteins, so one possibility is that inactivation of both HHP1 and the MBD proteins is necessary to fully disrupt the complexes. Although the MBD domain was identified based on its binding to methylated CG dinucleotides (^{m}CG) (Meehan et al., 1989), a previous study suggests that *Arabidopsis* MBD1, MBD2, and MBD4, unlike MBD5, MBD6, and MBD7, does not bind ^{m}CG (Zemach and Grafi, 2003). Proteins with MBD domains are frequently associated with histone deacetylases, both in animals and in plants (Nan et al., 1998; Ng et al., 1999; Zemach and Grafi, 2003). Thus, MBD1/2/4 may act as scaffold proteins involved in recruiting HDA6 and SANT1–SANT 4 rather than binding to ^{m}CG.

Preferential recruitment of domesticated *Harbinger* transposase to chromatin-modifying complexes

The two components of *Harbinger* transposases are typically domesticated as an interacting pair, as with ALP1/ALP2 and HDP1/HDP2 in plants or Harbi1/Naif in vertebrates (Sinzelle et al., 2008; Liang et al., 2015; Duan et al., 2017; Velanis et al., 2020). The sequence similarity between SANT1–SANT4 and the SANT/Myb/trihelix proteins of *Harbingers*, together with the ability of HHP1 and SANT2, SANT3, and SANT4 to interact directly, indicate that HHP1 and SANT1–SANT4 likely arose through co-domestication of ancestral *Harbinger*-encoded transposases. As observed previously (Duan et al., 2017; Velanis et al., 2020), the nuclease component tends to diverge less and is more easily identifiable than the SANT/myb/trihelix component of the complexes. Neither retain their original functions — HHP1 lacks the DDE catalytic triad necessary for nuclease activity, and SANT1–SANT4 lack the three tryptophan residues needed for DNA binding. Strikingly, in all three cases where domesticated *Harbinger* transposases have been functionally characterized in *Arabidopsis*, they appear to have been repurposed as components of different chromatin-modifying complexes: ALP1 and ALP2 are part of a Polycomb H3K27me3 histone methyltransferase complex, HDP1 and HDP2 are in a histone acetyltransferase complex, and HHP1 and SANT1–SANT4 in a histone deacetylase complex (Figure 6). One distinguishing feature of *Harbinger* TEs that may contribute to this status is their preferential insertion close to genes (Jiang et al., 2003; Zhang et al., 2004; Hancock et al., 2011), which may have favored the association of the host epigenetic machinery with *Harbinger* transposases, initially to recruit chromatin modifiers to the new TE and limit their effects on adjacent host gene activity.

The SANT complex influences histone deacetylation and flowering

The histone acetylation profiles and differentially expressed genes identified among the various genotypes tested here suggest that HDA6 and SANT proteins act together to repress common targets via histone deacetylation. Notably, the *sant-null* mutant and *hda6* mutant, and to a lesser extent the *hhp1* and *mbd1/2/4* mutants, affect gene expression similarly, whereas only the *hda6* mutant deregulates TE expression. We hypothesize that HDA6 mediates TE expression by participating in different complexes, such as those containing the H3K9me2 histone methyltransferases SUVH4, SUVH5, and SUVH6 (Yu et al., 2017). The SANT complex instead promotes flowering, by repressing the flowering repressors *FLC*, *MAF4*, and *MAF5*. Furthermore, GO term enrichment analysis reveals that a significant fraction of the genes that are differentially expressed in *sant-null* and *hda6*

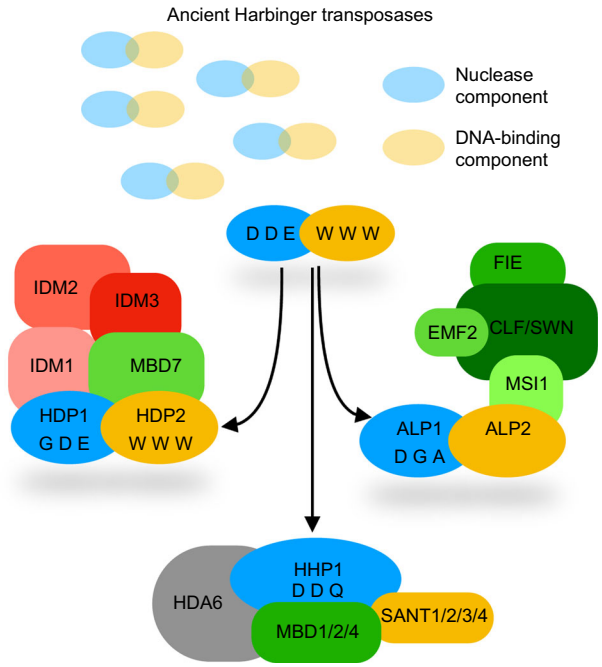

Figure 6. Harbinger transposases have been repurposed as accessory components of the epigenetic machinery

We propose that ancient Harbinger transposases have undergone domestication and functional diversification as epigenetic regulators. Nuclease component(s) then lost the catalytic DDE triad(s) and evolved into HDP1, ALP1, and HHP1. The DNA-binding partner(s), diverging to varying degrees, gave rise to HDP2, which retained DNA-binding activity, and to ALP2 and several SANT-Myb domain proteins that lost the WWW triad and presumably the ability to bind DNA. The physical association between the nuclease component derivative(s) and the DNA-binding component derivative(s) has been preserved, and the various heterodimers experienced neofunctionalization as components of distinct epigenetic complexes. The HDP1-HDP2 heterodimer was recruited by a HAT complex involved in DNA methylation. The ALP1-ALP2 heterodimer acquired function(s) antagonistic to PRC2, with which it associates physically, whereas the HHP1-SANTs heterodimer(s) were repurposed as accessories of novel HDA6-containing HDAC complex(es) (this study).

mutants take part in defense and abiotic stress responses, hinting that the novel histone deacetylase complex(es) might be crucial for coordinating a plethora of immunity and/or adaptive physiological responses.

MATERIALS AND METHODS

Plant materials and growth conditions

Arabidopsis thaliana accession Columbia-0 (Col-0) was used as the wild type. The T-DNA insertion lines *hhp1-1* (Salk_122829), *mbd1* (Salk_025352), *mbd4* (Salk_042834), and *sant3* (Salk_031353), as well as the EMS *hda6* allele (*axe1-5*), were obtained from the *Arabidopsis* Biological Resource Center. The *mbd2* mutant was generated via CRISPR/Cas9 genome editing, resulting in a single-nucleotide insertion in the first exon of *MBD2* (Figure S7B). The *mbd1/2/4* triple mutant was generated by genetic crossing of the *mbd1*, *mbd2*, and *mbd4* single mutants. The *hhp1-2* mutant, harboring a 16-bp deletion (Figure S7A), was also generated by CRISPR/Cas9-mediated editing, using the pHEE401 vector (plasmid #71286, Addgene) and a sgRNA with the sequence GUGUAGCUGUGGCGCUUAGG. The higher-order *sant* mutants, *sant-weak* and *sant-null*, both in the Salk_031353 background, were generated in the *sant3* (Salk_031353) mutant background via multiplex CRISPR/Cas9-mediated editing (Zhang et al., 2016) using a combination of three sgRNA pairs that targeted the four *SANT* genes (Figure S5). The primers used for genotyping and for sgRNA production are listed in Dataset S5.

The *35 S::HHP1-GFP* construct was generated by Gateway recombination between an entry clone containing the *HHP1* coding sequence and the *pGWB5* destination vector. The construct was introduced into Col-0 background by Agrobacterium (*Agrobacterium tumefaciens*)-mediated transformation via the floral dipping method. The complementation lines of *sant-null* were generated by cloning the *SANT1*, *SANT3*, and *SANT4* genomic coding regions and regulatory sequences into a modified pCAMBIA1305 vector, and subsequent Agrobacterium-mediated transformation of *sant-null* mutants via the floral dipping method. The primers used for plasmids construction are listed in Dataset S5.

Arabidopsis seedlings were grown on half-strength Murashige and Skoog (MS) medium under a long-day photoperiod (16 h light, 22°C/8 h darkness, 20°C), and then transferred to soil in growth chambers with the same conditions. To measure flowering time, the number of rosette leaves and number of days were recorded when the inflorescence stem reached 1 cm.

Immunoprecipitation and mass spectrometry (MS)

Flower tissue (3 g) collected from transgenic or wild-type plants was used for affinity purification. Tissue was ground in liquid nitrogen and homogenized in 15 mL of lysis buffer (50 mM Tris, pH 7.6, 150 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.1% NP-40, 0.5 mM dithiothreitol [DTT], 1 mM PMSF, protease inhibitor cocktail 1 tab/50 mL, Roche) for 10 min at 4°C. Following centrifugation, the supernatant was incubated with 10 μg of anti-Flag (Sigma, F1804) antibody or anti-GFP antibody (Invitrogen, A-11122) and 100 μL of Dynabeads Protein G (Invitrogen, 10003D) for 3 h at 4°C with agitation. Beads were then washed three times with lysis buffer and three times with wash buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$). Immunoprecipitated proteins were subjected to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis as previously described (Zhang et al., 2018).

Split luciferase assays

Split luciferase assays were carried out according to Lang et al. (Lang et al., 2015). The coding sequences of *HDA6*, *SANT1*, *SANT2*, *SANT3*, *SANT4*, *HHP1*, *MBD1*, *MBD2*, and *MBD4* were cloned into pCAMBIA-nLUC and pCAMBIA-cLUC vectors. Agrobacterium cultures (GV3101) harboring the indicated constructs were grown at 28°C for 24 h before collection by centrifugation. The cell pellets were then resuspended in infiltration

buffer (10 mM MES, pH 5.6, 10 mM $MgCl_2$, and 100 μM acetosyringone) to a final cell density of OD_{600} = 0.8. Equal volumes of resuspended cells were mixed in different combinations and incubated at room temperature for 2 h, before infiltration into the abaxial surface of *Nicotiana benthamiana* leaves. After infiltration, plants were returned to normal growth conditions for 48 h. Luciferase activity was detected with a luminescence imaging system (Princeton Instrument). The primers used for related plasmid construction are listed in Dataset S5.

Transcriptome deep sequencing (RNA-seq) and data analysis

We collected triplicate samples for the wild type, *hhp1-2*, *mbd1/2/4*, *sant-null*, and *hda6* for RNA-seq analysis. Total RNA was extracted from 12-day-old seedlings grown under long day conditions, with TRIzol reagent (Invitrogen) and then sent to Novogene Corporation (Beijing, China) for library construction and sequencing. Libraries were generated according to the manufacturer's instructions with the NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA) and sequenced on an Illumina NovaSeq. 6000 platform. After removing adapter sequences and low-quality reads, clean reads were mapped to the TAIR10 *Arabidopsis* genome using TopHat2 with the parameter "--b2-very-sensitive" (Kim et al., 2013). Mapped reads were visualized with Integrative Genomics Viewer (IGV) from the BAM files. Differential expression of genes and transposable elements was determined using Cuffdiff ($P < 0.05$; $|\log_2$ (fold change)$| > 1$) within Cufflinks (Trapnell et al., 2013). Heatmaps and boxplots of differentially expressed genes and TEs were generated in R. Gene ontology (GO) enrichment analysis was performed using the R package *clusterProfiler* in Bioconductor (Yu et al., 2012). The raw reads were deposited at the Gene Expression Omnibus (GEO) database (accession number: GSE167288).

Analyses of transcript levels by RT-qPCR

Total RNA from 12-day-old *Arabidopsis* seedlings grown on half-strength MS medium was extracted with TRIzol reagent (Invitrogen) and treated with DNase I (Takara) to remove contaminating genomic DNA. Total RNA was then reverse-transcribed into cDNA with PrimeScript RT reagent Kit (Takara, RR047A). Real-time quantitative PCR was performed on an Applied Biosystems 7500 Real-Time PCR System using ChamQ Universal SYBR qPCR Master Mix (Vazyme). *ACTIN7* was used as reference. Three biological replicates were performed for real-time PCR. The primers sequences can be found in Dataset S5.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) and data analysis

Chromatin immunoprecipitation (ChIP) was performed as previously described, with minor modifications (Zhu et al., 2012). Briefly, 2 g of 12-day-old seedlings grown on half-strength MS medium was collected and fixed in 1% formaldehyde for 15 min and then ground into powder in liquid nitrogen. Nuclei were isolated and chromatin was sheared into 200- to 500-bp fragments by sonication with a Bioruptor (Diagenode). After centrifugation, the sonicated chromatin was incubated with anti-acetylated Histone H3 antibody (Merck Millipore, 06-599) overnight and then incubated with Dynabeads Protein G (Invitrogen, 10003D) for 2 h with agitation at 4°C. The precipitated chromatin was washed and eluted with elution buffer (0.5% SDS and 0.1 M $NaHCO_3$) at room temperature, then concentrated via phenol–chloroform extraction and ethanol precipitation.

Immunoprecipitated DNA from three biological replicates were sent to Novogene Corporation (Beijing, China) for library construction and sequencing (Illumina NovaSeq. 6000; 150-bp pair-end reads). Adapter sequences and low-quality reads were removed from the raw data, after which clean reads were mapped to the TAIR10 *Arabidopsis* genome using Bowtie2 with default parameters; PCR duplicates were removed by using the Sambamba program (Langmead and Salzberg, 2012; Tarasov et al., 2015). Enriched ChIP peaks were identified with the MACS2 program and differentially enriched peaks were determined by the R package *DiffBind* in Bioconductor (FDR < 0.05; $\log_2$ (fold change) > 1). The read count of each gene region was calculated with the "intersect" command from the BEDTools suite (Quinlan, 2014). The H3KAc level of each gene is given as reads per kilobase per million (RPKM). The heatmap and boxplots of H3 acetylation levels in upregulated genes were generated in R. The raw H3Ac ChIP-seq data were deposited at the GEO database (accession number: GSE167288).

ACKNOWLEDGEMENTS

We thank the *Arabidopsis* Biological Resource Center for providing *Arabidopsis* mutant lines used in this study. This work was supported by grants from the National Key Research and Development Program of China (2020YFE0202300), the Central Public-interest Scientific Institution Basal Research Fund, the BBSRC under the Grant Reference BB/P008569/1 to J.G. C.N.V. and E.dL., and an Erasmus plus training award to L.G.

AUTHOR CONTRIBUTIONS

C.-J.Z., J.G., and J.-K.Z. designed the experiments; X.Z., J.H., C. N.V., Y.Z., Y.H., K.T., M.Z., L.G., E.dL., X.W., and L.Z. performed the experiments; X.Z., C.N.V., W.A.T., J.G., J.-K.Z., and C.-J.Z. analyzed the data; C.-J.Z., C.N.V., J.G., and J.-K.Z. wrote the article. All authors read and approved of the manuscript.

Edited by: Zhizhong Gong, China Agricultural University, China

Received Apr. 13, 2021; **Accepted** May 1, 2021; **Published** May 6, 2021

OO: OnlineOpen

REFERENCES

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. **21:** 381–395.

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvak, Z., Levin, H.L., Macfarlan, T.S., Mager, D.L., and Feschotte, C. (2018). Ten things you should know about transposable elements. Genome Biol. **19:** 199.

Bundock, P., and Hooykaas, P. (2005). An *Arabidopsis* hAT-like transposase is essential for plant development. Nature **436:** 282–284.

Chen, L.T., Luo, M., Wang, Y.Y., and Wu, K. (2010). Involvement of *Arabidopsis* histone deacetylase HDA6 in ABA and salt stress response. J. Exp. Bot. **61:** 3345–3353.

Chen, X., Ding, A.B., and Zhong, X. (2020). Functions and mechanisms of plant histone deacetylases. Sci. China: Life Sci. **63:** 206–216.

Cosby, R.L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E.J., and Feschotte, C. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. Science **371:** eabc6405.

Duan, C.G., Wang, X., Xie, S., Pan, L., Miki, D., Tang, K., Hsu, C.C., Lei, M., Zhong, Y., Hou, Y.J., Wang, Z., Zhang, Z., Mangrauthia, S.K., Xu, H., Zhang, H., Dilkes, B., Tao, W.A., and Zhu, J.K. (2017). A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation. Cell Res. **27:** 226–240.

Earley, K.W., Pontvianne, F., Wierzbicki, A.T., Blevins, T., Tucker, S., Costa-Nunes, P., Pontes, O., and Pikaard, C.S. (2010). Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. Genes Dev. **24:** 1119–1132.

Gu, X., Jiang, D., Yang, W., Jacob, Y., Michaels, S.D., and He, Y. (2011). *Arabidopsis* homologs of retinoblastoma-associated protein 46/48 associate with a histone deacetylase to act redundantly in chromatin silencing. PLoS Genet. **7:** e1002366.

Guo, J., Wei, L., Chen, S.S., Cai, X.W., Su, Y.N., Li, L., Chen, S., and He, X.J. (2021). The CBP/p300 histone acetyltransferases function as plant-specific MEDIATOR subunits in *Arabidopsis*. J. Integr. Plant Biol. **63:** 755–771.

Hancock, C.N., Zhang, F., Floyd, K., Richardson, A.O., Lafayette, P., Tucker, D., Wessler, S.R., and Parrott, W.A. (2011). The rice miniature inverted repeat transposable element mPing is an effective insertional mutagen in soybean. Plant Physiol. **157:** 552–562.

Hancock, C.N., Zhang, F., and Wessler, S.R. (2010). Transposition of the Tourist-MITE mPing in yeast: an assay that retains key features of catalysis by the class 2 PIF/Harbinger superfamily. Mob. DNA **1:** 5.

Henaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C., and Casacuberta, J.M. (2014). Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species. Plant J. **77:** 852–862.

Hollender, C., and Liu, Z. (2008). Histone deacetylase genes in *Arabidopsis* development. J. Integr. Plant Biol. **50:** 875–885.

Hosaka, A., and Kakutani, T. (2018). Transposable elements, genome evolution and transgenerational epigenetic variation. Curr. Opin. Genet. Dev. **49:** 43–48.

Hung, F.Y., Chen, F.F., Li, C., Chen, C., Lai, Y.C., Chen, J.H., Cui, Y., and Wu, K. (2018). The *Arabidopsis* LDL1/2-HDA6 histone modification complex is functionally associated with CCA1/LHY in regulation of circadian clock genes. Nucleic Acids Res. **46:** 10669–10681.

Jangam, D., Feschotte, C., and Betran, E. (2017). Transposable element domestication as an adaptation to evolutionary conflicts. Trends Genet. **33:** 817–831.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., and Wessler, S.R. (2003). An active DNA transposon family in rice. Nature **421:** 163–167.

Kapitonov, V.V., and Jurka, J. (2004). Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA Cell Biol. **23:** 311–324.

Kim, J.M., To, T.K., Matsui, A., Tanoi, K., Kobayashi, N.I., Matsuda, F., Habu, Y., Ogawa, D., Sakamoto, T., Matsunaga, S., Bashir, K., Rasheed, S., Ando, M., Takeda, H., Kawaura, K., Kusano, M., Fukushima, A., Endo, T.A., Kuromori, T., Ishida, J., Morosawa, T., Tanaka, M., Torii, C., Takebayashi, Y., Sakakibara, H., Ogihara, Y., Saito, K., Shinozaki, K., Devoto, A., and Seki, M. (2017). Acetate-mediated novel survival strategy against drought in plants. Nat. Plants **3:** 17097.

Lang, Z., Lei, M., Wang, X., Tang, K., Miki, D., Zhang, H., Mangrauthia, S.K., Liu, W., Nie, W., Ma, G., Yan, J., Duan, C.G., Hsu, C.C., Wang, C., Tao, W.A., Gong, Z., and Zhu, J.K. (2015). The methyl-CpG-binding protein MBD7 facilitates active DNA demethylation to limit DNA hyper-methylation and transcriptional gene silencing. Mol. Cell **57:** 971–983.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9:** 357–359.

Liang, S.C., Hartwig, B., Perera, P., Mora-Garcia, S., de Leau, E., Thornton, H., de Lima Alves, F., Rappsilber, J., Yang, S., James, G. V., Schneeberger, K., Finnegan, E.J., Turck, F., and Goodrich, J. (2015). Kicking against the PRCs -A domesticated transposase antagonises silencing mediated by polycomb group proteins and is an accessory component of polycomb repressive complex 2. PLoS Genet. **11:** e1005660.

Lin, J., Hung, F.Y., Ye, C., Hong, L., Shih, Y.H., Wu, K., and Li, Q.Q. (2020). HDA6-dependent histone deacetylation regulates mRNA polyadenylation in *Arabidopsis*. Genome Res. **30:** 1407–1417.

Liu, X., Yu, C.W., Duan, J., Luo, M., Wang, K., Tian, G., Cui, Y., and Wu, K. (2012). HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in *Arabidopsis*. Plant Physiol. **158:** 119–129.

Makarevitch, I., Waters, A.J., West, P.T., Stitzer, M., Hirsch, C.N., Ross-Ibarra, J., and Springer, N.M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet. **11:** e1004915.

Marmorstein, R., and Zhou, M.M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. Cold Spring Harb. Perspect. Biol. **6:** a018762.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell **58:** 499–507.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393:** 386–389.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat. Genet. **23:** 58–61.

Ning, Y.Q., Chen, Q., Lin, R.N., Li, Y.Q., Li, L., Chen, S., and He, X.J. (2019). The HDA19 histone deacetylase complex is involved in the regulation of flowering time in a photoperiod-dependent manner. Plant J. **98:** 448–464.

Oliver, K.R., McComb, J.A., and Greene, W.K. (2013). Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol. Evol. **5:** 1886–1901.

Pandey, R., Muller, A., Napoli, C.A., Selinger, D.A., Pikaard, C.S., Richards, E.J., Bender, J., Mount, D.W., and Jorgensen, R.A. (2002). Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic Acids Res. **30:** 5036–5055.

Probst, A.V., Fagard, M., Proux, F., Mourrain, P., Boutet, S., Earley, K., Lawrence, R.J., Pikaard, C.S., Murfett, J., Furner, I., Vaucheret, H., and Scheid, O.M. (2004). *Arabidopsis* histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. Plant Cell **16:** 1021–1034.

Quinlan, A.R. (2014). BEDTools: The swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics **47:** 11 12 11–34.

Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb. Perspect. Biol. **6:** a018713.

Shim, S., Lee, H.G., Lee, H., and Seo, P.J. (2020). H3K36me2 is highly correlated with m(6) A modifications in plants. J. Integr. Plant Biol. **62:** 1455–1460.

Sinzelle, L., Izsvak, Z., and Ivics, Z. (2009). Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell. Mol. Life Sci. **66:** 1073–1093.

Sinzelle, L., Kapitonov, V.V., Grzela, D.P., Jursch, T., Jurka, J., Izsvak, Z., and Ivics, Z. (2008). Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. Proc. Natl. Acad. Sci. USA. **105:** 4715–4720.

Steinbach, Y., and Hennig, L. (2014). *Arabidopsis* MSI1 functions in photoperiodic flowering time control. Front. Plant Sci. **5:** 77.

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. Genes Dev. **12:** 599–606.

Tanaka, M., Kikuchi, A., and Kamada, H. (2008). The *Arabidopsis* histone deacetylases HDA6 and HDA19 contribute to the repression of embryonic properties after germination. Plant Physiol. **146:** 149–161.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics **31:** 2032–2034.

To, T.K., Kim, J.M., Matsui, A., Kurihara, Y., Morosawa, T., Ishida, J., Tanaka, M., Endo, T., Kakutani, T., Toyoda, T., Kimura, H., Yokoyama, S., Shinozaki, K., and Seki, M. (2011). *Arabidopsis* HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. PLoS Genet. **7:** e1002055.

Velanis, C.N., Perera, P., Thomson, B., de Leau, E., Liang, S.C., Hartwig, B., Forderer, A., Thornton, H., Arede, P., Chen, J., Webb, K.M., Gümüs, S., Jaeger, G.D., Page, C.A., Hancock, C.N., Spanos, C., Rappsilber, J., Voigt, P., Turck, F., Wellmer, F., and Goodrich, J. (2020). The domesticated transposase ALP2 mediates formation of a novel Polycomb protein complex by direct interaction with MSI1, a core subunit of Polycomb Repressive Complex 2 (PRC2). PLoS Genet. **16:** e1008681.

Wu, K., Zhang, L., Zhou, C., Yu, C.W., and Chaikam, V. (2008). HDA6 is required for jasmonate response, senescence and flowering in *Arabidopsis*. J. Exp. Bot. **59:** 225–234.

Yang, J., Yuan, L., Yen, M.R., Zheng, F., Ji, R., Peng, T., Gu, D., Yang, S., Cui, Y., Chen, P.Y., Wu, K., and Liu, X. (2020). SWI3B and HDA6 interact and are required for transposon silencing in *Arabidopsis*. Plant J. **102:** 809–822.

Yruela, I., Moreno-Yruela, C., and Olsen, C.A. (2021). Zn(2+)-Dependent histone deacetylases in plants: structure and evolution. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2020.12.011

Yu, C.W., Chang, K.Y., and Wu, K. (2016). Genome-wide analysis of gene regulatory networks of the FVE-HDA6-FLD complex in *Arabidopsis*. Front. Plant Sci. **7:** 555.

Yu, C.W., Liu, X., Luo, M., Chen, C., Lin, X., Tian, G., Lu, Q., Cui, Y., and Wu, K. (2011). Histone deacetylase6 interacts with flowering locus D and regulates flowering in *Arabidopsis*. Plant Physiol. **156:** 173–184.

Yu, C.W., Tai, R., Wang, S.C., Yang, P., Luo, M., Yang, S., Cheng, K., Wang, W.C., Cheng, Y.S., and Wu, K. (2017). Histone deacetylase6 acts in concert with histone methyltransferases SUVH4, SUVH5, and SUVH6 to regulate transposon silencing. Plant Cell **29:** 1970–1983.

Zemach, A., and Grafi, G. (2003). Characterization of *Arabidopsis thaliana* methyl-CpG-binding domain (MBD) proteins. Plant J. **34:** 565–572.

Zhang, C., Du, X., Tang, K., Yang, Z., Pan, L., Zhu, P., Luo, J., Jiang, Y., Zhang, H., Wan, H., Wang, X., Wu, F., Tao, W.A., He, X.J., Zhang, H., Bressan, R.A., Du, J., and Zhu, J.K. (2018). *Arabidopsis* AGDP1 links H3K9me2 to DNA methylation in heterochromatin. Nat. Commun. **9:** 4547.

Zhang, X., Jiang, N., Feschotte, C., and Wessler, S.R. (2004). PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. Genetics **166:** 971–986.

Zhang, Z., Mao, Y., Ha, S., Liu, W., Botella, J.R., and Zhu, J.K. (2016). A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in *Arabidopsis*. Plant Cell Rep. **35:** 1519–1533.

Zhao, T., Zhan, Z., and Jiang, D. (2019). Histone modifications and their regulatory roles in plant development and environmental memory. J. Genet. Genomics **46:** 467–476.

Zhu, J.Y., Sun, Y., and Wang, Z.Y. (2012). Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq). Methods Mol. Biol. **876:** 173–188.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.13108/suppinfo

Figure S1. Phenotypic and genetic analysis of *hhp1* mutants

(**A**) Representative photographs of 53-day-old plants grown in short days. (**B**) Representative photographs of 39-day-old plants grown in short days. (**C**) Summary of flowering time in short days, expressed as rosette leaf number. Different letters above bars indicate significant differences (two-tailed Student's *t*-test, $P < 0.05$). Data are shown as means ± standard error of the mean. Scale bars, 1 cm.

Figure S2. Exploration of the interaction between HHP1, HDA6, MBD1, MBD2, and MBD4, and SANT1, SANT2, SANT3, and SANT4, as determined by split luciferase assays

(**A**) Luciferase complementation imaging (LCI) assay showing that MBD2 interacts strongly with HHP1, MBD1, MBD2, MBD4, and HDA6 and weakly with SANT1–SANT3. (**B**) LCI assay showing that MBD4 interacts strongly with HHP1, MBD1, MBD2, MBD4, HDA6, and SANT3 and weakly with SANT1, SANT2, and SANT4. **(C–F)** LCI assays showing that SANT1, SANT2, and SANT4 do not interact with one another.

Figure S3. Domain structure and phylogenetic analysis of SANT proteins

(**A**) Domain structure of SANT proteins. The domain architecture of each protein was determined from the SMART (Simple Modular Architecture Research Tool) database. Magenta denotes low-complexity regions. (**B**) Phylogenetic tree of the domesticated transposases ALP2 and HDP2, SANT protein, and Harbinger transposase DNA binding proteins.

Figure S4. SANT proteins share similarity with Harbinger transposase DNA binding proteins

Alignment of protein sequences for SANT1, SANT2, SANT3, SANT4, and Harbinger transposase DNA binding proteins. The three tryptophan residues that are highly conserved among Harbinger DNA binding proteins are highlighted in red boxes.

Figure S5. Genotyping analysis of *sant* mutants

(**A**) Schematic diagrams of *SANT1*, *SANT2*, *SANT3*, and *SANT4* loci. Dark blue and black boxes represent untranslated regions and exons, respectively, and black lines indicate introns. The black and orange arrows indicate the locations of PCR primers used in genotyping analysis and sgRNAs used for multiplex CRISPR/Cas9 editing, respectively. The same two sgRNAs targeted both *SANT1* and *SANT2*, which share very high sequence similarity. Black triangle indicates T-DNA insertion site of Salk_031353. (**B**) RT-PCR analysis of *SANT3* gene expression in Col-0 and the *sant-weak* and *sant-null* mutants. *ACTIN7* was included as reference. (**C**) Genotyping analysis of *SANT1*, *SANT2*, and *SANT4* in Col-0, *sant-weak*, and *sant-null*. (**D**) Electropherogram of the *SANT3* gene region in Col-0, *sant-weak*, and *sant-null* targeted by genome editing. The red dashed box indicates the 1-nt insertion in *sant-null* compared to Col-0. (**E**) Relative transcript levels of *At1g50830*, *At1g71890*, and *At5g54095*, as determined by RT-qPCR in Col-0, the *sant-null* mutant, and transgenic plants carrying the *SANT1*, *SANT3*, or *SANT4* genes introduced in the *sant-null* background. Error bars indicate the standard deviations of three technical repeats.

Figure S6. Flowering phenotypes of Col-0 and *sant-weak* and *sant-null* mutants grown in long-day conditions

Days to bolting (left) and number of rosette leaves upon bolting (right) for Col-0, *sant-weak*, and *sant-null* plants grown in long-day conditions. At least

20 plants were scored for each line. The *y*-axes denote days to bolting (left) and rosette leaf number (right). Each dot represents one plant. Different letters indicate statistically significant differences (ANOVA; Tukey HSD's multiple range test, $P < 0.05$).

Figure S7. Genotyping analysis of *hhp1-2* single mutant and *mbd1/2/4* triple mutants

(A) Electropherogram of *HHP1* in Col-0 and the *hhp1-2* mutant over the region targeted by genome editing. (**B**) Electropherogram of *MBD2* in Col-0 and the *mbd1/2/4* mutant over the region targeted by genome editing. The red dashed box indicates the 1-bp insertion in the mutant compared to wild type. **(C)** Schematic diagrams of the *MBD1* and *MBD4* loci indicating the positions of T-DNA insertion lines. Gray and black boxes represent untranslated regions and exons, respectively, and black lines indicate introns. The arrows indicate PCR primers used for genotyping. **(D)** Genotyping analysis of *mbd1* and *mbd4* T-DNA mutant alleles.

Figure S8. Venn diagrams showing the overlap of up-regulated (*P*-value<0.05 and a twofold cutoff was used) genes identified in *hda6*, *hhp1-2*, *mbd1/2/4* and *sant-null* The *P*-values were calculated using the hypergeometric distribution

Figure S9. Transcriptome analysis of genes upregulated in the *sant-null* mutant

(**A**) Heatmap representation of normalized expression in Col-0 and the indicated mutants of genes that were upregulated in *sant-null*. The color scale indicates normalized FPKM values. (**B**) Boxplot showing the expression in the indicated mutants of genes that were upregulated in *sant-null*. Each box represents $\log_2$(FPKM + 1 values of mutants/FPKM + 1 values of Col-0).

Figure S10. Gene Ontology (GO) term enrichment analysis of upregulated genes in *hda6* (left) and *sant-null* (right) The top 20 most enriched GO biological processes are shown as a bubble chart. The gene number (count) and *p*.adjust are given.

Figure S11. Heatmap representation

(**A**) and boxplot (**B**) of H3KAc levels for regions with higher levels of histone H3 acetylation in *sant-null* (left) and *hda6* (right) The color scale indicates normalized FPKM values.

Figure S12. The Histone H3 acetylation levels at *FLC*, *MAF4*, and *MAF5* as determined by ChIP-seq

Snapshots of the genome browser and bar graphs illustrate H3KAc levels at *FLC*, *MAF4*, and *MAF5* in Col-0, *sant-null*, and *hda6*. The results are from three replicates.

Dataset 1. RNA transcript levels of differentially expressed protein coding genes in *hhp1-2*

Dataset 2. RNA transcript levels of differentially expressed TEs in *hhp1-2*

Dataset 3. List of higher levels of H3Ac peaks in *sant-null* mutant

Dataset 4. H3KAc levels of *hda6*-mediated up-regulated genes

Dataset 5. List of primers used in this study