Univerzita Palackého v Olomouci

Přírodovědecká fakulta

# EPIGENETICKÁ KONTROLA ORGANIZACE CHROMATINU A JEJÍ VLIV NA REGULACI GENOVÉ EXPRESE ROSTLIN

Habilitační práce

**Mgr. habil. Aleš Pečinka, Ph.D.**

Olomouc 2022

**Prohlášení**

Prohlašuji, že jsem tuto habilitační práci zpracoval samostatně s použitím uvedených literárních zdrojů.

V Olomouci dne: 28. 01. 2022                                  Mgr. habil Aleš Pečinka, Ph.D.

# Obsah

# Přehled zkratek

**Všeobecné:**

| | |
|---|---|
| 5mC | 5-metylcytosin |
| 6mA | N6-metyladenin |
| bp | Bázové páry (délková jednotka DNA) |
| C | C-hodnota charakterizuje počet kopií genomu v buněčném jádře, 1C odpovídá obsahu DNA v jádře v haploidním nereplikovaném stavu. |
| DNA | Deoxyribonukleová kyselina |
| dsRNA | Dvouřetězcová RNA |
| Gbp | Giga bázové páry ($10^9$) |
| GBM | DNA metylace genů (angl. Gene body methylation) |
| H | Souhrné označení pro báze adenin, thymin a cytosin |
| HRE | Teplotně responsivní element (podle anglického Heat responsive element) |
| kbp | Tisíce bázových párů ($10^3$) |
| LTR | Dlouhé terminální repetice (podle anglického Long terminal repeat) |
| Mbp | Miliony bázových párů ($10^6$) |
| NOR | Oblast organizátorů jadérka (podle anglického Nucleolar organizer region) |
| nt | Nukleotidy (délková jednotka RNA) |
| RdDM | RNA řízená DNA metylace (podle anglického RNA-directed DNA Methylation) |
| rDNA | Ribozomální DNA locus |
| RNA | Ribonukleová kyselina |
| siRNA | Malá interferenční RNA (podle anglického small interfering RNA) |
| TE | Transponovatelný element |
| TIR | Terminální invertovaná repetice |
| TSD | Duplikace cílového místa (podle anglického Target site duplication) |
| WGD | Celogenomová duplikace (podle anglického Whole-genome duplication) |

**Jména genů, proteinů a komplexů**

| | |
|---|---|
| *AGO* | *ARGONAUTE*, podjednotka RNA Induced Silencing Complex (RISC) |
| *APE1L* | DNA-(apurinic or apyrimidinic site) 1 lyáza |
| *CENH3* | *CENTROMERIC HISTONE H3* |
| *CENP-A* | *CENTROMERIC PROTEIN-A*, viz. *CENH3* |
| *CMT* (*2* or *3*) | *CHROMOMETHYLASE* (*2* or *3*), DNA metyltransferáza |
| *DCL* | *DICER-LIKE*, RNáza H |
| *DDM1* | *DECREASED IN DNA METHYLATION 1*, chromatinový remodeler |
| *DME* | *DEMETER*, DNA glykosyláza |
| *DML1* | *DEMETER-LIKE 1*, DNA glykosyláza, syn. *ROS1* |
| *DRD1* | *DEFECTIVE IN RNA DIRECTED DNA METHYLATION 1*, chromatinový remodeler |
| *DRM* (*1* or *2*) | *DOMAINS REARRANGED METHYLTRANSFERASE* (*1* or *2*), *de novo* DNA metyltransferáza |
| *HEN1* | *HUA ENHANCER 1*, RNA metyltransferáza |
| *HOG1* | *S-ADENOSYL HOMOCYSTEINE HYDROLASE GENE 1* |
| *HSFA2* | *HEAT SHOCK FACTOR A 2*, teplotně responsivní transkripční faktor |
| *HSP70* | *HEAT SHOCK PROTEIN 70* |
| *KYP* | *KRYPTONITE*, histon H3 lysin 9 di-metyltransferáza |

| | |
|---|---|
| *LIG1* | *DNA LIGASE 1*, ligáza jednořetězcových zlomů DNA |
| *MET1* | *DNA METHYLTRANSFERASE 1*, CG DNA metyltransferáza |
| *MTHFD1* | *METHYLENETETRAHYDROFOLATE DEHYDROGENASE 1* |
| *NRPD1* | *NUCLEAR RNA POLYMERASE D 1*, největší podjednotka PolIV |
| PolII | DNA dependent RNA polymerase II |
| PolIII | DNA dependent RNA polymerase III |
| PolIV | DNA dependent RNA polymerase IV |
| PolV | DNA dependent RNA polymerase V |
| PRC (1 or 2) | Polycomb repressive complex (1 or 2) |
| *RDR (2 or 6)* | *RNA DEPENDENT RNA POLYMERASE (2 or 6)* |
| RISC | RNA-induced silencing complex |
| *ROS1* | *REPRESSOR OF SILENCING 1*; viz. *DML* |
| *VIM* | *VARIANT IN METHYLATION* |
| *ZDP* | DNA fosfatáza |

# 1. ZÁKLADNÍ CHARAKTERISTIKA JADERNÉHO GENOMU ROSTLIN

Deoxyribonukleová kyselina (DNA) je hlavním nositelem dědičné informace u všech živých organismů a některých virů (Craig et al., 2014). DNA je tvořena dvěma komplementárními řetězci, které vytváří dvoušroubicovou strukturu (Watson and Crick, 1953). Každý z řetězců je pak tvořen nukleotidy, které sestávají z nukleobáze - adeninu (A), guaninu (G), cytosinu (C) nebo tyminu (T), deoxyribosového cukru a fosfátové skupiny. Nukleotidové řetězce mohou být velmi dlouhé. Například délka 5,1 Giga párů bází (Gbp) jaderného genomu ječmene setého je rozděleno do sedmi DNA molekul – chromosomů – jejichž délka v rozvinutém stavu odpovídá přibližně 170 cm DNA (Mascher et al., 2017). Oba DNA řetězce drží pohromadě díky vodíkovým můstkům mezi komplementárními bázemi. Báze A a T mají dva vodíkové můstky, zatímco báze C a G mají můstky tři. Dvoušroubovice DNA se otáčí okolo své vlastní osy a vytváří jednu otáčku přibližně každých 10,4 bp. Replikací je DNA je přepisována do nových molekul DNA a transkripcí do ribonukleových kyselin (RNA), které pak slouží k translaci do proteinů, nebo mohou mít regulatorní funkci.

Celková DNA organismu – genom – je organizována do chromozomů (Kellogg and Bennetzen, 2004; Craig et al., 2014). Chromozomy prokaryot (baktérie a archaea) jsou kruhové a jsou uloženy přímo v cytoplazmě. Převážná část genetické informace eukaryot (protisté, rostliny, houby a živočichové) je lineární a je uložená v buněčném jádře, jehož obsah je oddělen od cytoplazmy jadernou membránou (Thanbichler and Shapiro, 2006). Malá část dědičné informace (prokaryotického původu a charakteru) je přítomna v mitochondriích všech eukaryot, a také v chloroplastech zelených rostlin. Počet chromozomů v buněčném jádře se může výrazně lišit. U rostlin je to od čtyř (2n = 4) u australské hvězdnicovité rostliny *Brachyscome dichromosomatica*, až po několik set u určitých stromů či primitivních kapradin (Khandelwal, 1990; Leach et al., 1995). Dále byl u rostlin zjištěn největší rozsah ve velikosti jaderného genomu ze všech eukaryontních skupin, a to od 63.6 Mbp/1C u bublinatky *Genlisea aurea*, až po 149 000 Mbp/1C u vraního oka *Paris japonica*, což odpovídá přibližně 2 365-násobnému rozdílu (Pellicer et al., 2014). Hodnota 1C je množství DNA organismu v haploidním nereplikovaném jádře. Jak velikost genomu tak i počet chromozomů jsou plastické znaky, které se mohou během evoluce poměrně rychle měnit (Johnston et al., 2005; Paterson et al., 2005; **Vu et al., 2015 - v této práci jsem se podílel na analýze rozdílných velikostí genomů u příbuzných druhů rodu *Genlisea***). Hlavním mechanismem vedoucím ke skokové změně v množství jaderné DNA jsou duplikace celého genomu (anglicky whole genome duplication, WGD). Celogenomové duplikace jsou časté u rostlin a v některých skupin živočichů. K WGD dochází buď zdvojnásobením chromozomů v rámci jednoho druhu

(autopolyploidizace), nebo hybridizací a splynutím genomů dvou druhů (alopolyploidizace). Semenné rostliny prodělaly během své evoluční historie nejméně jednu, častěji však více WGD událostí (Li et al., 2015b). WGD mají enormní vliv na evoluci genomu. Další kopie genomu snižují riziko negativního efektu v případě mutace jedné z kopií. WGD jsou považovány za urychlovače evoluce, které vedou k novým fenotypovým i ekologickým vlastnostem (Comai, 2005). Evoluční historie čeledí *Brassicaceae*, *Asteraceae*, *Fabaceae* nebo *Poaceae* ukazují, že změny spojené s polyploidizací pravděpodobně usnadňují vznik evolučních inovací a rychlou speciaci (Schranz et al., 2012). Evoluční význam polyploidizace je pravděpodobně zvýrazněn selekčním tlakem prostředí (Fawcett et al., 2009; De Smet et al., 2013). Po celogenomové duplikaci obvykle následuje postupná restrukturalizace genomu v řádu miliónů let, která zahrnuje přeskupení chromozomů řadou inverzí a translokací, postupnou ztrátu redundantních sekvencí a funkční diploidizaci (Lysak et al., 2006; Mandáková et al., 2010; De Smet et al., 2013). Tento cyklus se pak může opakovat v měřítku geologických dob.

Obrovské rozpětí velikostí rostlinných genomů je v ostrém kontrastu s relativně malou variabilitou v počtu protein kódujících genů. Například, huseníček rolní ($2n = 2x = 10$) s genomem $1C = 119$ Mbp má přibližně 27 000 genů, zatímco 142-krát větší genom hexaploidní pšenice seté ($2n = 6x = 42$, $1C = 17$ Gbp) jich obsahuje „jen" asi čtyřnásobek, tj. 108000 (Arabidopsis Genome Initiative, 2000; Appels et al., 2018). Navíc, po odečtení vlivu polyploidie má pšenice 36000 genů na subgenom, což je jen 1,33-krát více než huseníček.

Tento jev je nazýván paradoxem C-hodnoty (angl. C-value paradox) (Thomas, 1971). Na základě analýzy různých eukaryotických genomů je paradox C-hodnoty vysvětlován přítomností variabilního množství opakujících se (repetitivních) DNA sekvencí (Gregory, 2005; Tenaillon et al., 2010). Faktory určující množství repetitivní DNA nebo obecně velikost genomu zůstávají u mnoha druhů nejasné a nazývá se záhadou C-hodnoty (angl. C-value enigma) (Gregory, 2005).

## 2. HLAVNÍ OBLASTI JADERNÉHO GENOMU

Geny, mezigenové oblasti a repetitivní DNA představují hlavní oblasti jaderného genomu. Geny jsou základními funkčními jednotkami genomu. Zatímco počet chromosomů i stuktura genomu se v čase mohou měnit, některé geny zůstávají obdivuhodně stabilní a jejich původ lze vysledovat již do doby před rozdělením hlavních eukaryotických skupin. Primárním produktem genu je RNA transkript, který slouží jako návod pro tvorbu specifického proteinu nebo může mít regulační funkci (van Driel et al., 2003). Geny jsou řízeny cis-regulačními elementy, které jsou v rostlinných genomch umístěny obvykle v mezigenovém prostoru před místem aktivace transkripce, vzácněji také v intronech či za místem transkripční terminace. Mezigenové prostory tedy hrají důležitou roli nejen ve fyzické separaci jednotlivých genů, ale také v jejich regulaci. U huseníčku rolního představují genové oblasti (včetně exonů a intronů) 50,4% jaderného a repetitivní DNA 19,1% (Arabidopsis Genome Initiative, 2000). Pro srovnání, u velkého a repeticemi bohatého genomu ječmene setého tvoří geny 2% a repetitivní DNA 80% genomu (Wicker et al., 2017). Zatímco funkce genů a jejich regulačních oblastí je zřejmá, u repetitivní DNA je tomu naopak. Repetice byly dlouho považovány za jakousi zbytnou část genomu bez zjevného užitku. Přestože toto bezpochyby platí pro mnoho kopií různých repetic, existuje dostatek důkazů o tom, že repetitivní DNA je nutná pro správnou funkci rostlinných genomů (Lisch, 2013). Repetitivní DNA lze rozdělit na dvě hlavních skupiny – tandemové a roztroušené repetice.

### 2.1 Tandemové repetice – opomíjená, ale nepostradatelná součást genomu

Tandemové repetice (nazývané také jako satelitní DNA) jsou uspořádány do dlouhých řad sestávajících z několika až tisíců kopií s velmi podobnou nebo identickou nukleotidovou sekvencí. Mezi tandemové repetice patří také základní strukturní oblasti eukaryotického chromosomu, kterými jsou centromerické, telomerické a ribozomální DNA (rDNA) oblasti.

Centromera je přítomna prakticky u všech eukaryotických chromozomů a je nutná pro správné dělení sesterských chromatid do dceřiných buněk (Przewloka and Glover, 2009). Oblast centromerických repetic tvoří obvykle několik Mbp dlouhé pole kratších (<500 bp) tandemových repetic, které se rychle vyvíjejí, a jsou často druhově specifické (Melters et al., 2013). Funkční centromera je pak definována v rámci tohoto pole repeticemi, které nesou centromerickou variantu histonu H3 nazývanou CENP-A nebo CENH3 (Lermontova et al., 2014). Tato oblast slouží jako platforma pro vazbu kinetochoru.

Telomerické repetice obvykle vytvářejí kratší (<100 kbp) pole na koncích chromozomů a mají obecně konzervovanější sekvenci než centromery. Telomery obratlovců a některých

rostlin se skládají z opakující se sekvence TTAGGG. U většiny rostlin (včetně huseníčku rolního nebo obilovin) jsou však telomery tvořeny sekvencí TTTAGGG (Richards and Ausubel, 1988; Watson and Riha, 2010). Několik málo rostlinných skupin obsahuje jiné typy telomerických repetic, jako je např. CTCGGTTATGGG u rodu *Allium* (Fajkus et al., 2016). Jak se tyto neobvyklé telomery vyvinuly, zůstává nejasné. Funkcí telomer je chránit konce chromosomu před degradací, a to vytvořením dlouhého jednořetězcového přesahu DNA, který vytváří oblouk, tzv. T-loop, a vmezeřuje se do telomerické dvouřetězcové DNA a vytváří menší, tzv. D-loop (Riha et al., 2006). Některé konce telomer, však mohou být ukončeny přímo dvouřetězcovým DNA zlomem a jejich ochrana před degradací je zprostředkována přítomností proteinového komplexu Ku70/Ku80 (Kazda et al., 2012). Telomery jsou prodlužovány enzymem telomerázou za pomocí specifické RNA molekuly, přičemž u rostlin byla tato RNA podjednotka telomerázy nalezena teprve v nedávno (Fajkus et al., 2019).

Třetí obligátní tandemové repetice zahrnují ribozomální DNA (rDNA), které kódují sekvence nezbytné pro sestavení RNA podjednotek ribozomů. rDNA se u rostlin vyskytují ve dvou hlavních typech 5S a 45S rDNA. Monomer 5S rDNA má délku 120 bp a je organizován do úseků dlouhých 100 až 200 kbp, obsahujících přibližně 800 až 1500 repetic. Struktura 45S rDNA jednotek je složitější. Každá jednotka (~9 kbp) má tři podjednotky 18S, 5.8S a 28S, které jsou odděleny interními transkribovanými spacery a vně pak 5 'a 3' externími transkribovanými spacery. Referenční genom huseníčku rolního Col-0 obsahuje dvě sady 45S rDNA, každou přibližně se 375 kopiemi, které se nachází v subtelomerických oblastech krátkých ramének chromozomů 2 a 4 (Copenhaver and Pikaard, 1996). Počet 45S a 5S rDNA kopií a lokusů se může se lišit i mezi populacemi stejného druhu (Roa and Guerra, 2012; Long et al., 2013). Aktivní kopie 45S rDNA produkují obrovské množství rRNA a spoluvytváří jádérko, které je po aplikaci barviv vážících se na DNA mikroskopicky pozorovatelné jako slabě obarvená oblast. Proto je 45S rDNA také označována jako oblast organizátorů jadérka (NOR).

Dále obsahují rostlinné genomy řadu roztroušených tandemových repetic o různé délce a počtu kopií. Evoluce těchto repetic, jejich funkce a regulace zůstávají poměrně málo známé. Zdá se, že rychle vznikají a pravděpodobně i zanikají. U několika druhů bylo zjištěno velké množství kopií určitých repetic roztroušeně po celém genomu. Příkladem mohou být řebčíky (rod *Fritillaria*), řepa (rod *Beta*) nebo některé brukvovité (Ambrožová et al., 2011; Zakrzewski et al., 2014; **Finke et al., 2019 - zde jsme nalezli specifickou tandemovou repetici, která tvoří více než 10% genomu u Australské brukvovité rostliny *Ballantinia antipoda*. Tato repetice je DNA hypometylovaná a výrazně ovlivňuje strukturu genomu**). U všech těchto případů je pozoruhodné, že abundatní tandemové repetice jsou sekvenčně velmi bohaté na A a T báze (až 80%). To naznačuje, že AT-bohaté repetice jsou hostitelským genomem obtížně

odstranitelné nebo epigeneticky kontrolovatelné. Zda existuje příčinná souvislost mezi A-T bohatostí, počtem kopií a jejich epigenetickou kontrolou tandemových repetic je však nejasné.

## 2.2 Transponovatelné elementy – genomičtí parazité mnoha funkcí

Roztroušené repetice jsou heterogenní skupinou, která zahrnuje všechny opakovaně se vyskytující nekódující sekvence. Nejlépe popsanými roztroušenými repeticemi jsou transponovatelné elementy (transposony, TE). Na rozdíl od protein kódujících genů a tandemových repetic mohou transpozony aktivně měnit svou pozici v hostitelském genomu. Transpozice do genu může vyvolat ztrátu nebo zisk funkce genu a tím fenotypovou změnu. Klasickým příkladem jsou transpozicí vyvolané změny v pigmentaci osemení kukuřice, které vedly popsání transpozonů jako „skákajících genů" Barbarou McClintock ve 40. letech 20. století (McClintock, 1950). Za tento revoluční objev jí byla v roce 1983 udělena Nobelova cena za fyziologii nebo medicínu (https://www.nobelprize.org/prizes/medicine/1983/summary/). Transpozony jsou tradičně rozděleny do dvou hlavních tříd. Třída I obsahuje retrotranspozony, které „skáčí" pomocí RNA meziproduktu mechanismem „zkopíruj se a vlož se" (angl. copy-and-paste). Třída II pak obsahuje transpozony, které se transponují bez RNA meziproduktu mechanismem „vystřihni se a vlož se" (angl. cut-and-paste). Transpozony nesoucí všechny části nutné pro transpozici se nazývají autonomní, zatímco ty, které využívají molekulární komponenty odvozené od jiných elementů jsou neautonomní.

Retrotranspozony zahrnují nejméně šest podtříd, které se liší svou strukturou a organizací protein kódující části: *LONG TERMINAL REPEAT* (*LTR*), *DICTYOSTELIUM INTERMEDIATE REPEAT* (*DIR*), *PENELOPE-LIKE ELEMENTS* (*PLEs*), *LONG INTERSPERSED ELEMENTS* (*LINEs*), *SHORT INTERSPERSED ELEMENTS* (*SINEs*) a *SADHU* (Rangwala et al., 2006; Wicker et al., 2007). *DIR* nebyly u rostlin nalezeny a *PLEs* jsou vzácné (Arkhipova, 2006). Naopak elementy všech ostatních skupin jsou u rostlin časté. *COPIA* a *Gypsy LTR* retrotranspozony dokonce tvoří podstatnou část mnoha rostlinných genomů (Feschotte et al., 2002; Wicker et al., 2017; Appels et al., 2018). Kódující část úplných *LTR* retrotranspozonů je na obou stranách ohraničena *LTR* oblastmi, které slouží jako *cis*-regulační oblasti transpozonu (Casacuberta and Santiago, 2003). *LTR* na 5' konci přímo navazuje na tzv. vazebné místo primeru (primer binding site), kde je zahájena reverzní transkripce. Kódující úsek retrotranspozonu pak končí polypurinovým traktem, jehož úkolem je zabránit štěpení transkriptu RNázou H, a také slouží k zahájení syntézy druhého řetězce (Havecker et al., 2004). Kódující oblast retrotranspozonu zahrnuje: částice podobné viru (virus-like particles), Reverzní transkriptázu, Integrázu, Aspartovou proteázu a RNázu H. Další početnou skupinou autonomomních elementů jsou *LINE*, které však postrádají *LTR*. Oproti

tomu, *SINE* a *SADHU* jsou malé neautonomní traspozony, které se mobilizují pomocí proteinů produkovaných *LINE* elementy (Rangwala and Richards, 2010).

Transpozony třídy II reprezentují podtřídy DNA transpozon, *HELITRON* a *MAVERIC*. Nukleáza DNA transpozonů indukuje během inzerce dvouřetězcový zlom cílové sekvence, zatímco nukleázy podtříd *HELITRON* a *MAVERIC* indukují pouze jednořetězcový zlom. DNA transpozony se pak dále dělí na dvě hlavní skupiny: *TERMINAL INVERTED REPEAT* (*TIR*) transpozony a pro houby specifickou skupinu *CRYPTON*. V místě inzerce *TIR* elementu vzniká typická stopa tzv. target site duplication (TSD). Všech devět rodin *TIR* elementů pak lze rozlišit na základě jejich specifických *TIR* a *TSD* sekvencí. U podtřídy *HELITRON* se vyvinul zvláštní druh amplifikace, který je založen na tzv. „rolling-circle" replikačním mechanismu (Yang and Bennetzen, 2009). *HELITRONy* jsou běžné v mnoha rostlinných genomech (Xiong et al., 2014).

Primární aktivitou transpozonů je jejich vlastní množení v hostitelském genomu. Rostliny a jiné organismy však vyvinuly specifické mechanismy, které vedou k potlačení jak transkripční tak transpoziční aktivity transpozonů. Tyto mechanismy jsou podrobně popsány v následujících kapitolách. U řady transpozonů naopak došlo k jakési „domestikaci" v rámci daného genomu, kdy se daný transpozon stává funkčním elementem zapojeným do např.: (i) regulace genové transkripce, (ii) duplikování protein kódujících genů nebo (iii) tvorbě genetické variability.

Geny v blízkosti transpozonů jsou obvykle transkribovány slaběji (Hollister et al., 2011) a transpozony mohou dokonce představovat hlavní *cis*-regulační element genu (Kinoshita et al., 2007; Butelli et al., 2012; **Pietzenuk et al., 2016 - v rámci této práce jsme ukázali, že inzerce teplotně responzivního transpozonu do blízkosti genu způsobila jeho aktivovatelnost teplotním stresem**). Transposony a další repetitivní sekvence hrají nepostradatelnou roli ve vytváření sekvenční a strukturní variability v rámci rostlinných genomů (Springer et al., 2016; **Pecinka et al., 2013 - v této přehledné práci jsme shrnuli vliv různých transposonů na regulaci genové exprese a základní funkce genomu**; Lisch, 2013). Méně známo je, že transpozony mohou vytvářet duplikace genů mechanismem retropozice. Tento proces zahrnuje reverzní transkripci genové mRNA pomocí retrotranspozonové (obvykle *LINE*) reverzní transkriptázy a její integraci do hostitelského genomu (Kaessmann et al., 2009). Výsledná duplikovaná kopie pak obvykle neobsahuje introny a může postrádat funkční promotor. Na rozdíl od duplikací genů na bázi DNA má retropozice vysoký potenciál generovat evoluční inovace ve velmi krátké době, např. expresí retrokopií v novém vývojovém kontextu, generováním chimérických genů s novými kombinacemi domén nebo v rámci horizontálního přenosu genů (Wang et al., 2006a; Yoshida et al., 2010; Sakai et al., 2011). Přibližně 1% všech genů v rámci rostliných genomů vzniklo retropozicí (Zhang et al., 2005; Wang et al., 2006a,

2006b; **Abdelsamad and Pecinka, 2014 - zde jsme identifikovali dosud nejvyšší známý počet retrogenů v genomu huseníčku rolního a ukázali, že řada retrogenů je transktripčně aktivována v pylových zrnech**).
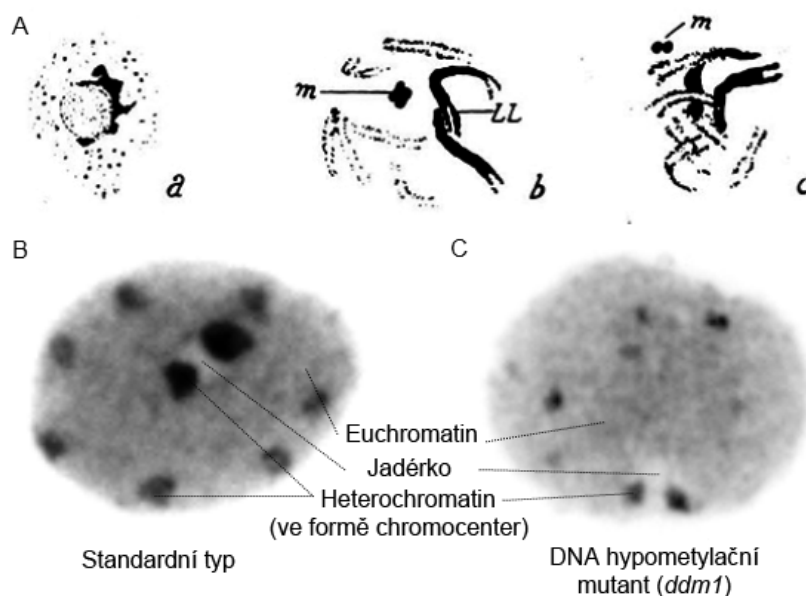
# 3. ORGANIZACE EUCHROMATINU A HETEROCHROMATINU V BUNĚČNÝCH JÁDRECH ROSTLIN

DNA je v buněčném jádře uložena ve formě chromatinu, tj. komplexu DNA a s ní asociovaných proteinů (Li et al., 2007). Mezi hlavní funkce chromatinu patří: (i) úsporné a přitom funkční uložení DNA v omezeném prostoru buněčného jádra, (ii) příprava chromozomů k buněčnému dělení, (iii) potlačením aktivity transpozonů a virů a tím ochrana stability DNA a (iv) kontrola transkripce a replikace (Kouzarides, 2007; Alabert and Groth, 2012; Jeppsson et al., 2014; Chen et al., 2021).

Nejčastějšími chromatinovými proteiny jsou histony. Tyto vysoce alkalické proteiny mají silnou vazebnou afinitu k DNA. Komplex osmi histonových molekul, tj. dvě molekuly každého z histonů H2A, H2B, H3 a H4, asociovaných k DNA vytváří základní jednotku chromatinu – nukleosom. Během tvorby nukleosomu se nejprve vytvoří dva dimery H3 a H4, které se spojí v tetramer. Následně se přidají dva H2A-H2B dimery za vzniku histonového oktameru, okolo kterého se omotá přibližně 147 bp (1,67 otočky) DNA. U některých nukleosomů poté ještě dojde k zafixování pomocí linkerových histonů H1 nebo H5. Navinutí DNA okolo histonového oktameru výrazně napomáhá kompaktnímu uložení DNA v buněčném jádře, ale zároveň nechává DNA relativně přístupnou pro další procesy. To je klíčové s ohledem na množství dědičné informace. Například jádro lidské buňky o průměru pouhých 6 μm obsahuje přibližně 2 metry DNA. To odpovídá asi 40 km vlákna smotaného do tenisového míčku (Alberts, 2002). Podobné poměry lze očekávat také pro rostlinná jádra. Řetězec nukleosomů na DNA tvoří chromatinové vlákno o průměru asi 10 nm a představuje relativně otevřenou, transkripčně permisivní, strukturu. Naproti tomu nahloučené nukleosomy vytvářejí kompaktní a transkripčně represivní chromatin.

Detailní analýza struktury a funkce chromatinu byla umožněna od 80. letech 20. století nástupem řady nových metod genetiky, molekulární biologie a biochemie, jako jsou např. genetické screeny, PCR, fluorescenční *in situ* hybridizace, imunolokalizace, fluorescenční mikroskopie, blotovací techniky, analýzy methylace DNA, chromatinová imunoprecipitace apod. V posledních dvou desetiletích byla analýza chromatinu a transkripce rozšířena na celý genom s využitím různých typů mikročipů (microarrays) a vysoce výkonných metod hlubokého sekvenování (Redman et al., 2004; Rehrauer et al., 2010; Reuter et al., 2015), které představují skutečnou revoluci v přístupu ke studiu organizace jádra a regulace genové exprese. Nově se pak přidaly metody analýzy jednotlivých buněk (angl. single cell genomics), které umožňují mnohem přesnější analýzu s ohledem na přítomnost či absenci určitých chromatinových modifikací, otevřenost chromatinu apod. (Dorrity et al., 2021; Chanou and Hamperl, 2021).

Různé formy chromatinu byly popvé popsány německo-švýcarským botanikem a genetikem Emilem Heitzem ve 20. letech 20. století u játrovky *Pellia epiphylla*. Heitz pozoroval slabě a silně nabarvené oblasti chromatinu na mitotických chromosomech a dokonce i v interfázních jádrech (Heitz, 1928). Tyto domény následně pojmenoval jako „euchromatin" a „heterochromatin" (**Obrázek č. 1**). Oba termíny se staly populárními, protože dobře vystihují dva hlavním chromatinové stavy existující v eukaryotických jádrech: euchromatin je rozvolněný, bohatý na geny a obvykle transkripčně aktivní. Naopak, heterochromatin je kondenzovaný, obvykle transkripčně neaktivní a s vysokým podílem repetitivních sekvencí (Bártová et al., 2008). Struktura a distribuce euchromatinu a heterochromatinu v genomu závisí na mnoha faktorech jako jsou např. taxonomická skupina, množství a rozložení repetitivní DNA na chromosomech nebo typ buněk (Ernst et al., 2011; Tiang et al., 2012). Struktura chromatinu se dále výrazně mění v průběhu buněčného cyklu. Zatímco v interfázních jádrech má chromatin poměrně otevřenou strukturu, tak na počátku buněčného dělění dochází k jeho shlukování do kompaktních struktur vyšších řádů, které vytvářejí mitotické a meiotické chromozomy s mikroskopicky pozorovatelnými sesterskými chromatidami. Na konci buněčného dělění dochází opět k rozvolnění.



**Obrázek č. 1. Euchromatin a heterochromatin.** (A) Nákresy interfázního jádra (a) a mitotických chromosomů (b,c) játrovky *Pellia epiphylla*. Převzato z (Heitz, 1928). Oblasti heterochromatinu jsou černě a euchromatinu světle. (B,C) Interfázní jádra z listů huseníčku rolního – divokého typu (B) a DNA hypo-metylačního mutanta *ddm1* (C). Euchromatin se jeví šedě a hetero-chromatin vytváří tmavá chromocentra, která lokalizují po obvodu jádra nebo sousedí s jadérkem (slabě barvená oblast). Množství heterochromatinu je epigeneticky kontrolovaný proces. Příkladem může být mutace v genu chromatin remodelačního faktoru *DDM1*, která způsobuje výraznou redukci množství heterochromatinu a chromocenter.

Přestože je rozložení euchromatinu a heterochromatinu v rostlinném jádře často druhově specifické, lze vysledovat minimálně jeden obecný trend a to lokalizaci heterochromatinu na periferii jádra. U obratlovců, kde se jednotlivé chromozomy výrazně liší co do množství transposonů a obsahu heterochromatinu, vykazují celé chromozomy preferenčně centrální nebo periferní lokalizaci. Tato organizace je pak velmi výrazná u ptáků, jejichž četné malé a na geny bohaté chromosomy leží v centru jádra, zatímco velké genově chudé chromozomy vytvářejí „vrstvu heterochromatinu" na vnějším obvodu jádra (Habermann et al., 2001). Podobně striktní radiální organizace dosud nebyla u rostlin pozorována, což je pravděpodobně způsobeno relativně vysokou mírou uniformity mezi chromosomy v rámci jednoho genomu co do poměru AT a CG bazí a rodin repetitivních elementů. Rostliny s malými genomy (<1 Gbp) a nízkým obsahem repetitivní DNA mají obvykle heterochromatin nahloučený v poměrně malé části chromosomu, která tvoří během interfáze tzv. chromocentra (**Obrázek 1B**). Chromocentra jsou lokalizována na jaderné periferii a z nich vystupují chromosomová ramena vytvářející tzv. chromosomová teritoria (Pecinka et al., 2004). U huseníčku rolního a písečného (*Arabidopsis thaliana* a *A. arenosa*) jsou jednotlivé chromosomy rozmístěny v jádře náhodně, vyjímku však tvoří chromosomy nesoucí 45S rDNA, které jsou v častém kontaktu s jadérkem i sebou navzájem (Berr et al., 2006; Pecinka et al., 2004). U rostlin s velkými genomy jako je ječmen setý nebo pšenice setá zaujímají interfázní chromozomy tzv. Rabl-orientaci, kdy centromery a telomery vytváří shluky na protilehlých pólech jádra (Jasencakova et al., 2001; Doğan and Liu, 2018). Z toho vyvozujeme, že chromosomy v Rabl konfiguraci zaujímají „tvar písmene V". Tato organizace odráží uspořádání chromozomů na konci buněčného dělení a zároveň může sloužit jako příprava pro další dělení. Hranice mezi euchromatinem a heterochromatinem je u druhů s velkými genomy méně ostrá, ale lze rozlišit heterochromatický a euchromatický pól buněčného jádra (Fuchs et al., 2006). To je dáno tím, že euchromatické protein kódující geny jsou nahloučeny na koncích chromosomů a v Rabl organizaci tak dochází ke koncentraci euchromatinu na telomerovém pólu jádra. Zda je Rabl organizace čistě výsledkem množství jaderné DNA, mitotické aktivity buněk, jejich kombinací, či jiných faktorů dosud není známo. Je třeba zdůraznit, že mnoho druhů vykazuje přechodné stavy mezi non-Rabl a Rabl chromozomovou organizací.

Důvody pro umístění heterochromatinových oblastí na periferii buněčného jádra zůstávají relativně neznámé, a proto uvádím několik, vzájemně se nevylučujících, možností týkající se možných benefitů této organizace. (i) Vazba centromerických repetic na jadernou membránu může zjednodušovat vazbu mikrotubulů na kinetochory. (ii) Periferní oblasti jádra mohou být častěji vystaveny škodlivým faktorům jako jsou reaktivními formy kyslíku nebo UV-B záření. Heterochromatin by tak mohl sloužit jako jakýsi ochranný „štít" s nízkou hustotou

protein kódujících genů. (iii) Akumulace mutací v repetitivní DNA (ad ii) by dokonce mohla napomáhat inaktivaci transpozonů a očistě genomu od těchto genomických parazitů (**Willing et al., 2016 - v této studii jsme vystavovali huseníček rolní působení simulovaného slunečního záření a ukázali jsme, že mutace vznikají především v pozici metylovaných cytosinů v repetitivních oblastech genomu**). (iv) Fyzická separace heterochromatinu jednotlivých chromosomů může bránit mechanismu homologní rekombinace mezi sekvenčně téměř identickými avšak nehomologními úseky chromosomů a tím snižovat riziko vzniku dicentrických nebo acentrických chromozomů (Chiolo et al., 2011). Tento model však nemusí platit pro druhy s Rabl organizací, kde jsou centromery nahloučeny relativně blízko u sebe, resp. by vyžadoval jejich relokalizaci směrem do euchromatické části jádra.

## 3.1 Epigenetický základ euchromatinu a heterochromatinu

Na molekulární úrovni jsou euchromatin a heterochromatin definovány epigeneticky. Termín „epigenetika" (latinsky epi = nad) zavedl v roce 1942 britský biolog Conrad Hall Waddington, aby popsal diferenciaci lidských kmenových buněk (Waddington, 1942). Od devadesátých let 20. století se termín epigenetika používá k popisu „stabilně dědičného fenotypu vyplývajícího ze změn v chromatinu, které však nejsou změnami DNA sekvence" (Berger et al., 2009). V současnosti je termín epigenetika v literatuře často používán v širším významu, který zahrnuje téměř jakékoliv (včetně krátkodobých) změn. Tyto změny, které nejsou přenášeny mitotickým dělením je dle mého názoru lepší popisovat jako dynamiku a funkce chromatinu a nikoliv jako epigenetické jevy *sensu stricto* (**Pecinka and Mittelsten Scheid, 2012 - v tomto přehledném článku jsme shrnuli aktuální znalosti ohledně transkripční regulace genové exprese u rostlin a navrhli standardy pro posuzování výsledků epigenetických experimentů, především s ohledem na stresem indukované a mezigenerační efekty**).

### *3.1.1 DNA metylace jako klíčová epigenetická modifikace rostlin*

Metyl skupina (-CH$_3$) představuje důležitou epigenetickou značku (Ratel et al., 2006; Nabel et al., 2012). U DNA rostlin je nalézána ve formě 5-metylcytosinu (5mC) nebo N6-metyladeninu (6mA; **Obrázek 2**). 6mA byl sice detekován u řady organismů, u rostlin však představuje pouze asi 0,5% všech adeninů, což naznačuje, že se jedná o relativně vzácnou modifikaci (Vanyushin et al., 1988; Fu et al., 2015; Luo et al., 2015). 6mA se pravděpodobně podílí na polohování nukleosomů a regulaci transkripce (Fu et al., 2015).

**Obrázek č. 2. Chemická struktura metylovaných nukleobazí.**

Cytosin      5-metylcytosin      Adenin      N6-metyladenin
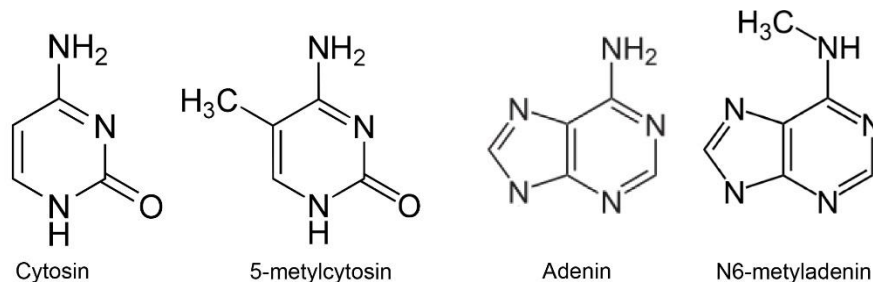
Nejčastěji studovanou modifikací DNA je 5mC (dále jen jako DNA methylace). Přítomnost a význam 5mC se liší v závislosti na fylogenetické skupině. DNA metylace je vzácná u hmyzu a mnoha hub, je však častá u savců a rostlin (Feng et al., 2010; Takayama et al., 2014). U savců dochází k metylaci DNA *de novo* během raného embryonálního vývoje a pouze cytosiny následované guaniny (CG kontext) zůstávají metylovány v somatických buňkách. Rostliny mají komplexní systém, kdy dochází k DNA metylaci ve třech funkčně odlišných sekvenčních kontextech: CG, CHG a CHH (kde H je C, A nebo T) a tato metylace přetrvává relativně stabilně v průběhu celého vývoje. DNA metylace vykazuje specifické rozmístění v rostlinných genomech. Je obohacena v heterochromatických a redukována v euchromatických oblastech (**Obrázek 3A, B**).

Celogenomové studie u huseníčku rolního odhalily, že přibližně 20% genů nese DNA metylaci také v genech (anglicky tzv. gene body methylation, GBM). Oproti repetitivním sekvencím, které mají methylované cytosiny ve všech sekvenčních kontextech, se GBM vyskytuje výhradně v CG kontextu (**Obrázek 3C**) (Zilberman et al., 2008). Přítomnost GBM je charakteristickým znakem především dlouhých, evolučně konzervovaných a stabilně přepisovaných genů (Coleman-Derr and Zilberman, 2012a; Takuno and Gaut, 2012). V současné době je stále diskutována funkce této metylace např. při prevenci transkripce z kryptických promotorů anebo zprostředkovávání alternativního sestřihu (Takuno and Gaut, 2012). Analýza DNA metylačních profilů u více než 1 000 přírodních linií huseníčku rolního z různých geografických oblastí severní polokoule potvrdila pozitivní korelaci mezi GBM a vysokou transkripcí, a navíc ukázala, že je tato metylace nižší v regionech s teplejším podnebím (Kawakatsu et al., 2016). Nově bylo zjištěno, že GBM zcela chybí u několika druhů z čeledi brukvovitých (Bewick et al., 2017; Bewick and Schmitz, 2017). To naznačuje, že tato metylace může mít regulatorní funkci, ale není pro genom rostlin esenciální.

Nejasná role CG metylace genů kontrastuje s rozhodující úlohou DNA metylace pro stanovení konstitutivního heterochromatinu u rostlin. V heterochromatinu jsou cytosiny metylovány ve třech kontextech CG, CHG a CHH (**Obrázek 3C**). Tato akumulace metylace vede k potlačení transkripce prostřednictvím řady kroků, které zahrnují instalaci represivních

chromatinových značek, odstranění permisivních značek a zahuštění chromatinu (Law and Jacobsen, 2010; Fultz et al., 2015). Ztráta schopnosti založit a udržet heterochromatinovou struktutu určitých částí genomu vede u rostlin k vývojovým poruchám, silně snížené kondici a fertilitě, či dokonce k letalitě (Miura et al., 2001; Mathieu et al., 2007; Mirouze et al., 2009). Proto bude epigenetická kontrola heterochromatické metylace DNA podrobně popsána v následujících kapitolách.
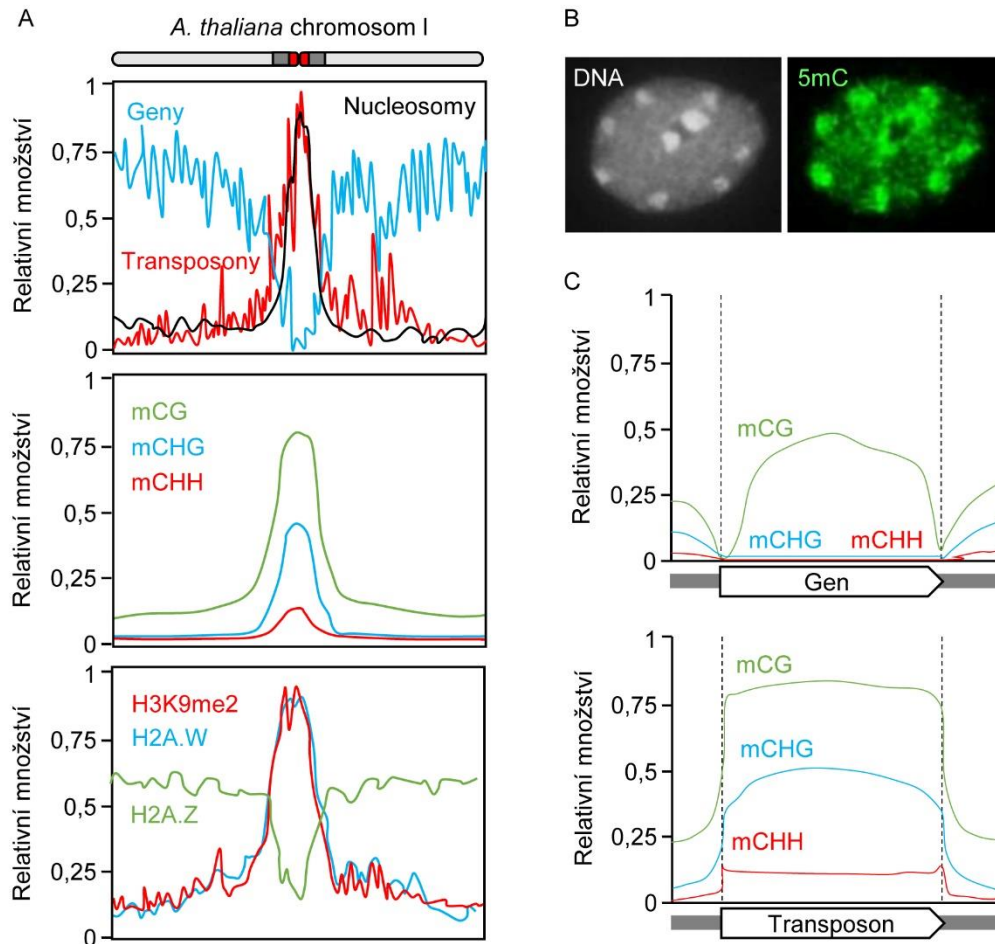


**Obrázek č. 3. Rozmístění DNA metylace u huseníčku rolního (*Arabidopsis thaliana*).** (**A**) Shora dolů: Schematické zobrazení chromosomu 1 (ca. 30 Mbp). Centromera je vyznačena červeně, pericentromerický heterochromatin tmavě šedě a euchromatická chromozomová ramena světle šedě. x-osa grafů odpovídá chromosomu 1 a y-osa pak ukazuje relativní frekvenci. Horní graf ukazuje pozitivní korelaci mezi přítomností transpozonů (TE) a hustotou nukleosomů a negativní korelaci s distribucí genů. Střední graf shrnuje rozmístění DNA metylace v jednotlivých sekvenčních kontextech. Spodní graf pak shrnuje rozmístění heterochromatické modifikace H3K9me2 a varianty H2A.W a dále euchromatické varianty H2A.Z. Chromatinové profily byly převzaty z citovaných prací a upraveny (Bernatavichute et al., 2008; Chodavarapu et al., 2010; Coleman-Derr and Zilberman, 2012a; Yelagandula et al., 2014). (**B**) Interfázní jádro obarvené DAPI (vlevo) a protilátkou proti 5mC (vpravo), které výrazně barví oblasti heterochromatických chromocenter (na obrázku vpravo se jeví jako nejsvětlejší oblasti jádra). (**C**) Relativní frekvence DNA methylace v různých sekvenčních kontextech v genech a transpozonech. Převzato a upraveno z (Coleman-Derr and Zilberman, 2012a).

### 3.1.2 Mechanismus de novo metylace dosud nemetylovaných oblastí

Strukturované rozmístění DNA metylace v jaderném genomu naznačuje, že toto rozmístění není náhodné (Chan et al., 2005; **Pecinka et al., 2013 - v tomto přehledném článku jsem shrnul aktuální znalosti týkající se DNA metylace u rostlin a typické vlastnosti DNA metylovaných lokusů**). Přestože dosud nejsou podmínky, které vedou k DNA metylaci konkrétního lokusu zcela známy, bylo prokázáno, že ve zvýšené míře korelují s:

      (i) tandemovými i roztroušenými repeticemi

      (ii) obrácenými repeticemi, které mohou vytvářet tzv. vlásenkovou strukturu

      (iii) lokusy produkujícími nestandardní transkripty (anti-sense, nedokonale terminované apod.)

      (iv) geny řízenými silnými (často virovými) promotory

      (v) sekvenční homologií s již metylovanými lokusy.

DNA sekvence nově vkládané do genomu (např. během DNA transformace) jsou bez DNA methylace. V určitých případech je však rostlina metyluje *de novo*. V některých případech tak dochází k umlčení transgenních konstruktů vkládaných do rostlinné DNA a absenci požadované vlastnosti či dokonce vypnutí dalších konstruktů nesoucích homologní sekvence (Daxinger et al., 2008). Přestože se jedná o komplikaci v procesu produkce transgenních rostlin, byly tyto fenotypy klíčové pro mechanistické pochopení procesu instalace a udržení DNA metylace u rostlin. Tento proces je rozdělen do několika drah. Standardní *de novo* DNA metylační dráha je závislá na produkci malých interferenčních RNA molekul (angl. short interfering RNA, siRNA) s úplnou homologií k cílové sekvenci, která je následně metylována. Proto se tato dráha často nazývá jako RNA-řízená DNA metylace (RdDM) (**Obrázek 4**) (Matzke and Mosher, 2014; Fultz et al., 2015). Během RdDM je cílový lokus nejprve transkribován DNA-dependentními RNA polymerázami II a IV (PolII a PolIV), přičemž druhá zmíněná polymeráza se vyskytuje pouze u cévnatých rostlin a je specifická právě pro RdDM dráhu. Tyto transkripty musí vytvořit dvouklánovou vlásenkovou strukturu, a to buď jednoduchým přeložením (v případě přirozených obrácených repetic) nebo pomocí RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), která je schopná dosyntetizovat chybějící řetězec. RNA vlásenka je poté štěpena pomocí endoribonukleázy DICER-LIKE 3 (DCL3) na 24 bp dlouhé siRNA úseky. Tyto krátké dvouřetězcové RNA molekuly jsou stabilizovány methylací 3' konců pomocí RNA metyltransferázy HUA ENHANCER 1 (HEN1). Jeden řetězec siRNA pak tvoří RNA-indukovaný umlčovací komplex (RNA-induced silencing complex, RISC) s proteinem

ARGONAUTE 4 (AGO4). Následně se RISC spojí s komplexem RNA-dependentní polymerázy V (PolV), který zajišťuje DNA *de novo* metylační reakci. V tomto procesu je soustředěna řada enzymů a strukturních proteinů, z nichž uvádím pouze některé. DECREASED RNA-DEPENDENT DNA METHYLATION 1 (DRD1) je chromatinový remodeler specifický pro RdDM, který s největší pravděpodobností odemkne specifické nukleosomy a umožní tak přístup *de novo* DNA metyltransferázám z rodiny DOMAINS REARRANGED METHYLTRANSFERASE (DRM) k cílovým sekvencím.

Vedle kanonické *de novo* DNA metylace existuje také alternativní dráha, která je někdy nazývána jako post-transkripční genové vypínání (podle anglického Post-transcriptional gene silencing, PTGS) (**Obrázek 5**). V rámci této dráhy jsou jednořetězcové mRNA generované DNA-dependentní Polymerázou II převedeny na dvouřetězcové pomocí RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), naštěpeny DICER-LIKE 2 (DCL2) nebo DICER-LIKE 4 (DCL4) na 21 až 22 bp dlouhé dsRNA, jejichž jeden řetězec je navázán do ARGONAUTE 1 (AGO1) a vzniklý komplex štěpí mRNA nebo může (v omezené míře) vést k *de novo* methylaci DNA (Matzke and Mosher, 2014). RdDM je velmi účinný mechanismus, který metyluje nově vložené repetitivní sekvence, transpozony nebo cizí sekvence a to v některých případech již v první generaci (Marí-ordóñez et al., 2013). Tento jev byl poprvé mechanisticky popsán u tzv. kosuprese, kdy transformace petúnií transgenem, který měl způsobit výraznější zabarvení okvětí, vedla naopak úplné ztrátě exprese a bílým květům (Van Blokland et al., 1994).

RdDM aktivita může být pozorována takřka v přímém přenosu během fenoménu nazývaného paramutace. U tohoto epigenetického jevu, přenáší epigeneticky umlčená alela svůj reprimovaný expresní status na geneticky identickou, avšak exprimovanou alelu a způsobí její vypnutí. Reprimovaná epialela je tedy dominantní a tuto vlastnost přenáší také na původně aktivní epialely, které tímto umlčuje (Chandler and Stam, 2004; Chandler and Alleman, 2008; Pilu, 2015). K vypnutí dochází pomocí *de novo* DNA metylace, která je naváděna k naivní exprimované epialele pomocí *trans* aktivních siRNA molekul produkovaných reprimovanou alelou. Paramutace se výrazně uplatňuje v epigenetické regulaci u některých druhů rostlin a typickým příkladem je kukuřice, kde byla popsána řada klasických případů (Arteaga-Vazquez and Chandler, 2010)

Přes svou složitost a zapojení velkého množství faktorů má vyřazení RdDM aktivity ve standardních podmínkách překvapivě malý vliv na fenotyp (u huseníčku obvykle pouze mírně opožděné kvetení). Množství DNA metylace zůstává téměř beze změny a rostliny vykazují relativně malé změny v expresi genů a transpozonů (Huettel et al., 2006; Stroud et al., 2012a; Zemach et al., 2013). To naznačuje, že RdDM funguje alespoň částečně redundantně s jinou molekulární drahou.
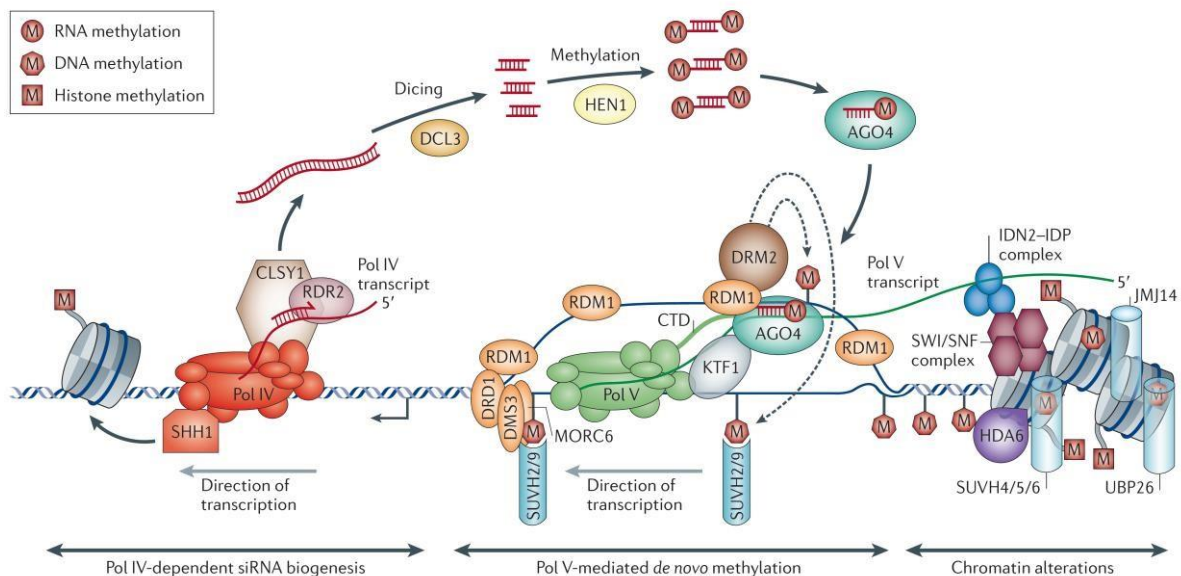
**Obrázek 4. Model standardní** *de novo* **RNA-dependentní DNA metylační dráhy (převzato z Matzke and Mosher, 2014).** Jednotlivé kroky v rámci této dráhy jsou popsány v textu.
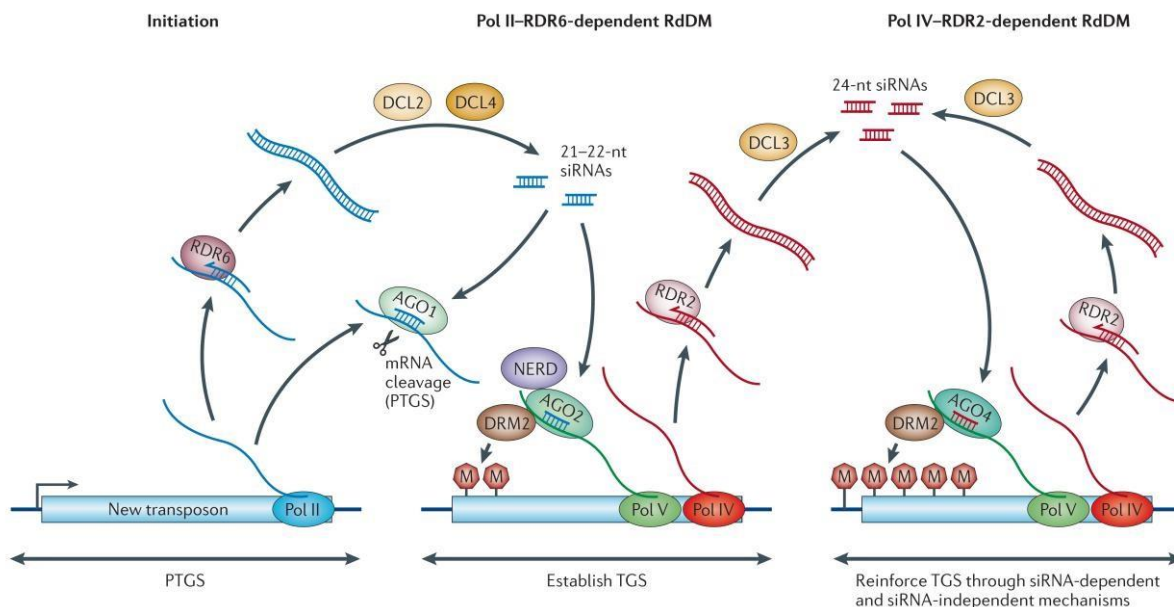


**Obrázek 5. Model alternativní** *de novo* **DNA metylační dráhy (převzato z Matzke and Mosher, 2014).** Jednotlivé kroky této dráhy jsou popsány v textu.

### 3.1.3 Udržovací DNA metylační dráha zajišťuje stabilní hladinu DNA metylace

U huseníčku rolního byla provedena řada dopředných genetických screenů zaměřených na identifikaci klíčových genů pro kontrolu DNA metylace a transkripční represi repetitivní DNA.

Tyto screeny odhalily skupinu mutantů charakterizovaných obvykle výrazným snížením DNA metylace a masivním navýšením transkripční a transpoziční aktivity transpozonů.

První skupina genů ovlivňovala DNA metylaci ve všech sekvenčních kontextech a obsahovala primárně enzymy nutné pro syntézu S-adenosyl methioninu, který je používán různými typy metyltransferáz jako donor metyl skupiny. Dva nejlépe charakterizované enzymy z této skupiny jsou ETHYLENETETRAHYDROFOLATE DEHYDROGENASE/ METHENYLTETRAHYDROFOLATE CYCLOHYDROLASE (MTHFD1), který se podílí na folátovém cyklu a S-ADENOSYL HOMOCYSTEINE HYDROLASE GENE 1 (HOG1), který je nutný pro zdárný průběh metioninového cyklu (Rocha et al., 2005; **Baubec et al., 2010 – v rámci této publikace jsme identifikovali a charakterizovali novou mutantní alelu genu HOG1**, Groth et al., 2016). Silné mutantní alely MTHFD1 a HOG1 jsou letální. Nicméně částečná ztráta funkce těchto genů má za následek globálně sníženou methylaci DNA i histonů, výrazně zhoršený růst rostlin a ztrátu umlčení transpozonů (**Baubec et al., 2010**, Groth et al., 2016). Předpokládá se, že hydrofolátový cyklus je pojítkem mezi výživou a úrovní DNA metylace u rostlin (Groth et al., 2016). Tento příklad upomíná na systém *Agouti* u myší (*Mus musculus*), kde strava bohatá na hydrofoláty vede k normální metylaci DNA a nízké expresi genu *Agouti*, což se projevuje tmavou srstí (Waterland and Jirtle, 2003). Na hydrofoláty chudá strava pak vede k nižší úrovni DNA metylace, expresi genu *Agouti* a následně nažloutlé srsti. Významným hráčem v globální methylaci DNA je faktor remodelace chromatinu DECREASED IN DNA METHYLATION 1 (DDM1). Předpokládalo se, že DDM1 řídí methylaci DNA tím, že umožňuje přístup DNA metyltransferáz k nukleosomální DNA uzamčené histonem H1 v heterochromatinových oblastech (Zemach et al., 2013). Nicméně, nejnovější výzkum ukazuje, že DDM1 je spíše zodpovědný za instalaci heterochromaticky specifické histonové varianty H2A.Z (Osakabe et al., 2021), jejíž přítomnost by pak vedla k reprimaci chromatinu a instalaci DNA metylace.

Dále bylo nalezeno několik faktorů ovlivňujících DNA metylaci v sekvenčně specifickém kontextu (**Obrázek 6**). V CG kontextu je metylace DNA udržována komplexem DNA METHYLTRANSFERASE 1 (MET1) a rodinou částečně redundantních proteinů obsahujících SRA a SET domény, které byly u huseníčku rolního popsány jako VARIANT IN METHYLATION (VIM) (Mathieu et al., 2007; Woo et al., 2007; Hashimoto et al., 2008; Woo et al., 2008). CG methylace se vyskytuje jak v kódujících oblastech genů, tak v transpozonech a je s největší pravděpodobností kopírována z metylovaného řetězce DNA na nově nasyntetizovaný řetězec během replikace DNA. Během tohoto procesu se VIM proteiny nejprve váží na molekulu DNA, invadují její centrální část a vytáčí cytosiny v místě CG dinukleotidů tak, aby mohla proběhnout metylace pomocí MET1 (Hashimoto et al., 2008). Ztráta aktivity

MET1 vede k závažným vývojovým defektům a podobně jako u *ddm1* mutantů také k mobilizaci transpozonů (Finnegan et al., 1996; Mathieu et al., 2007; Mirouze et al., 2009).
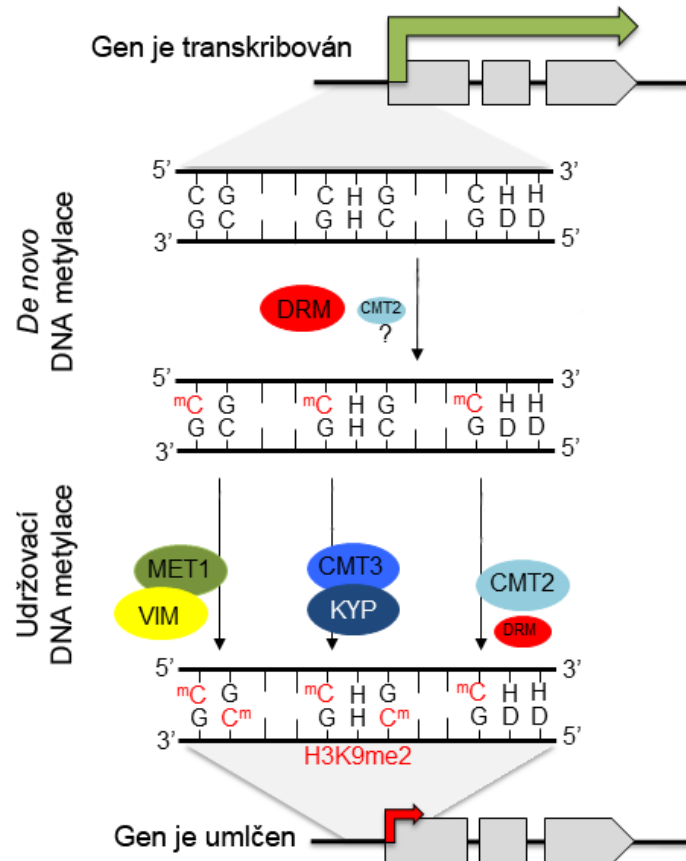


**Obrázek 6. Přehled DNA metylačních drah u huseníčku rolního.** Transkripčně aktivní gen (zelená šipka) může být *de novo* DNA metylován v CG, CHG and CHH contextu pomocí kanonické nebo alternativní RNA-řízené DNA Metylační (RdDM) dráhy. V procesu *de novo* DNA metylace jsou klíčové DRM DNA metyltransferázy (viz. **Obrázky 4** a **5** pro detailní informace) a potenciálně také CMT2. Tato první vlna DNA metylace slouží jako templát pro replikačně vázanou udržovací DNA metylaci. V CG kontextu je tato metylace instalována MET1 – VIM komplexem. V CHG kontextu je to CMT3, která interaguje s histonovou metyltransferázou KYP, což je podtrženo kolokalizací CHG DNA metylace a H3K9me2 metylace. CHH metylace na okrajích dlouhých transpozonů je udržována díky aktivitě CMT2 (a pravděpodobně RdDM jako záložní dráhy). Hypermetylace DNA ve všech třech funkčních kontextech vede k transkripčnímu vypnutí genu.

Kromě evolučně konzervované CG methylační dráhy mají rostliny také unikátní rodinu DNA methyltransferáz obsahujících chromodoménou tzv. CHROMOMETHYLÁZ (Bartee et al., 2001). První funkčně charakterizovaný enzym této rodiny byl CHROMOMETHYLASE 3 (CMT3), který metyluje specificky v CHG kontextu. CMT3 interaguje, prostřednictvím svých chromo- a „bromo adjacent homology" (BAH) domén, s histonovou methyltransferázou KRYPTONITE (KYP; syn. SET DOMAIN PROTEIN 33; syn. SU(VAR) 3-9 HOMOLOG 4), která di-metyluje histon H3 v pozici lysinu 9 (H3K9me2) a tak dále přispívá k tvorbě

heterochromatinu (Bartee et al., 2001; Bernatavichute et al., 2008; Du et al., 2012). Nedávno bylo zjištěno, že jiný protein této rodiny CHROMOMETHYLASE 2 (CMT2) zprostředkovává CHH DNA metylaci na okrajích dlouhých transpozonů (Zemach et al., 2013). Tím se liší od *de novo* DNA metyltransferázy DRM2, která metyluje primárně kratší transpozony nebo deleční zbytky dlouhých transpozonů, jako jsou například tzv. soloLTR (solitérní dlouhé terminální repetice LTR transpozonů) (Huettel et al., 2006). Faktory řídící odlišnou lokalizaci CMT2 a DRM2, zůstávají neznámé, ale je možné, že CMT2 je také naváděna určitými histonovými modifikacemi.

Ztráta asi 30% heterochromatinu, která byla pozorována u mutantů v *DDM1* nebo *MET1*, vede k mnoha pleiotropním fenotypům včetně mobilizace transpozonů, deregulaci stovek genů, nestabilitě genomu, vývojovým problémům a snížené fertilitě (Jeddeloh et al., 1999; Miura et al., 2001; Shaked et al., 2006; Mathieu et al., 2007; Mirouze et al., 2009; **Baubec et al., 2010 - v rámci této práce jsme izolovali a popsali fenotypy několika mutantních alel DDM1**). Mobilizace transpozonů a jejich inzerce do protein kódujících genů je vážnou hrozbou, protože přímo ovlivňuje funkčnost genů a je nejpravděpodobnějším faktorem způsobujícím redukci životnosti v rámci generací *ddm1* a *met1* mutantů (Miura et al., 2001; Mathieu et al., 2007; Mirouze et al., 2009). Ztráta DNA methylace navíc v některých případech vede k transkripci z normálně umlčených nebo vývojově regulovaných genů. Některé z genů objevených tímto způsobem byly důležité pro pochopení epigenetického základu specifických biologických procesů, jako je např. genový imprinting (Köhler et al., 2012; Batista and Köhler, 2020). Mutanti v genu *DDM1* také vykazují zvýšenou nestabilitu genomu a vyšší frekvenci meiotických rekombinací (Shaked et al., 2006; Melamed-Bessudo et al., 2005). To je s největší pravděpodobností způsobeno rozvolněnou strukturou chromatinu a/nebo stříháním DNA transpozonovými nukleázami. To poukazuje na důležitost represe transpozonů a tvorby heterochromatinu pro kontrolu genové transkripce a stability genomu rostlin.

V souhrnu, tato kapitola ilustruje složitost a také částečnou redundanci mechanismů methylace DNA u rostlin (**Obrázek 6**). Systém paralelních a záložních drah umožňuje přesnou regulaci genové exprese a především funkční ochranu rostlinného jaderného genomu před genomickými parazity typu transpozonů nebo virů. Dále je třeba zdůraznit, že v rostlinách stále existuje několik dalších necharakterizovaných DNA methyltransferáz.

### 3.1.4 Aktivní DNA demetylace u rostlinných pomocí bázově excizních oprav DNA

Přestože je heterochromatická DNA metylace považována za jednu z nejstabilnějších epigenetických značek, existují buňky či oblasti genomu, kde mohou být metylované báze cíleně odstraňovány (**Obrázek 7**). U rostlin je tento proces aktivní DNA demetylace prováděn rodinou DEMETER (DME) a DEMETER-LIKE (DML1, DML2 and DML3) bifunkčních DNA glykosyláz/lyáz, které odstraňují 5-metylcytosin v rámci dvouřetězcové DNA na principu bázových excisních oprav DNA (Penterman et al., 2007; Hsieh et al., 2009; Zhu, 2009).
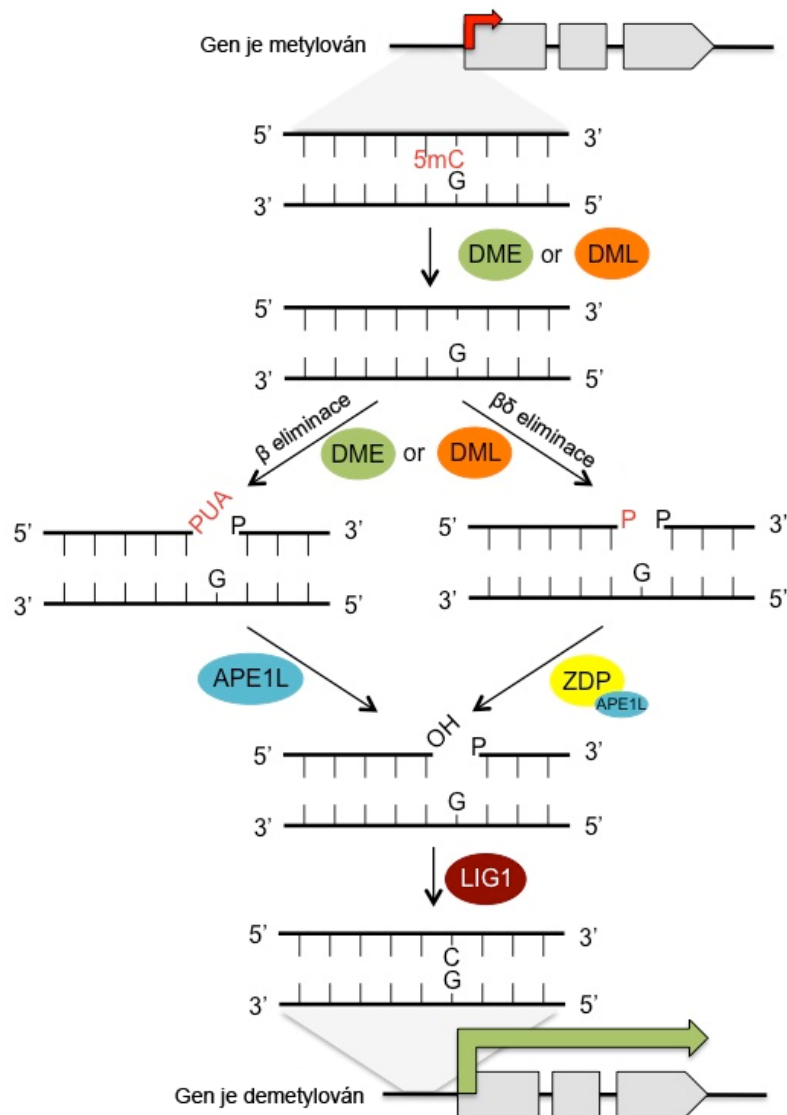


**Obrázek 7**. **Aktivní demetylace DNA.** Odstranění metylovaných bazí z regulační oblasti genu může vést k jeho aktivaci. Molekulární funkce jednotlivýc enzymů je popsána v textu. Obrázek byl převzat z (Li et al., 2015a) a upraven.

Tyto enzymy narušují DNA kostru v abazickém místě pomocí beta- nebo beta/delta-eliminace (Choi et al., 2002; Gong et al., 2002). To vytváří jednořetězcové poškození DNA charakterizované buď fosfo-a,b-nesaturovaným aldehydem (PUA) nebo fosfátovou skupinou na 3' konci, které musí být odstraněny DNA lyásou APE1L nebo DNA fosfatázou ZDP (Martínez-Macías et al., 2012; Lee et al., 2014; Li et al., 2015a). V konečném kroku dojde k uzavření léze nemetylovaným cytosinem pomocí DNA LIGÁZY 1 (LIG1) (Li et al., 2015a).

DNA demetylázy DME a DMLs fungují jako pozitivní regulátory genové transkripce, kdy např. odstraňují DNA metylaci z promotorů specifických genů a tak je mohou aktivovat. Nicméně způsob jakým jsou demetylázy naváděny na místa určení zůstává v současnosti neznámý. Zatímco DML1/ROS1 je aktivní především během somatického vývoje, tak DME je specializován na demetylaci v samičím gametofytu před oplozením.

### 3.1.5 Histonové varianty a modifikace

Histony jsou velmi důležitým nositelem epigenetické informace a jedním ze základních „kamenů" chromatinu. Jsou tvořeny C-terminální globulární doménou, která je situována ve středu nukleosomu, a nestrukturovanou N-terminální částí, která vystupuje z nukleosomomu do prostoru a slouží jako substrát pro posttranslační modifikace jakými jsou např. metylace (Me), acetylace (Ac), fosforylace (P), ubiquitinace (Ub) a citrulinace (Cit) (**Obrázek 8**).
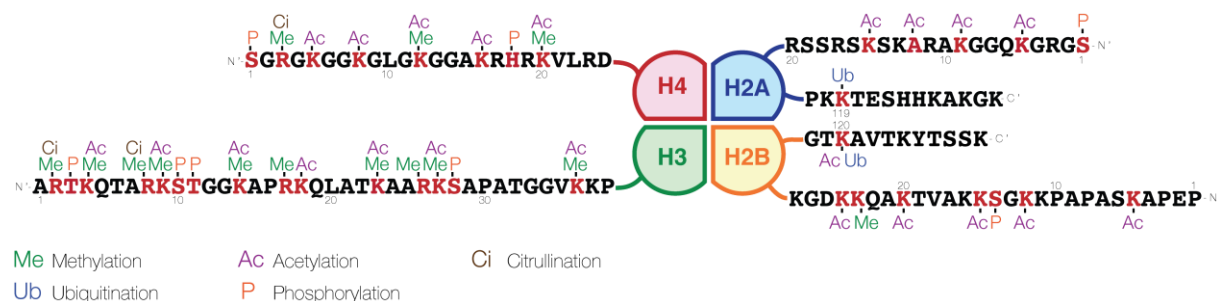


**Obrázek 8. Posttranslační modifikace histonových N-terminálních konců**. (Zdroj: Wikipedia).

V rámci této práce budou diskutovány pouze histonové metylace a acetylace, protože tyto modifikace slouží velmi dobře jako diagnostické značky euchromatinu a heterochromatinu. Metylovány mohou být lysiny (K) a argininy (R), a oproti jiným modifikacím se metylace může vyskytovat jako mono-, di- nebo trimetylace, přičemž každá modifikace má specifickou signální či regulatorní funkci. Například histon H3 lysin 9 di-metylace (H3K9me2) je transkripčně represivní heterochromatinová modifikace, zatímco H3K9me3 funguje u rostlin jako transkripčně permisivní euchromatinová modifikace (Roudier et al., 2011). Metylace histonů je katalyzována rodinou histonových metyltransferáz, které jsou charakterizovány

přítomností SET domény. Tato rodina je poměrně rozsáhlá a např. u huseníčku rolního je tvořena více než 35 geny. To naznačuje, že metylace histonů představuje poměrně komplexní systém, kdy jednotlivé geny mohou být jak pozičně tak pletivově specifické. Metylace histonů může být také odstraněna a tuto funkci zprostředkovávají histon demetylázy, které obsahují Jumonji C (jmjC) doménu a u rostlin stále zůstávají poměrně málo prozkoumanou skupinou enzymů (Liu et al., 2010; Qian et al., 2019). Podobně jako v případě histonové metylace, existují také enzymy, které zprostředkovávájí acetylaci histonů tzv. histon acetyltransferázy, resp. jejich deacetylaci, která je zajišťována histon deacetylázami. Acetylace histonů má obecně transkripčně permisivní funkci a obohacena v euchromatinu, zatímco deacetylace histonů transkripci tlumí.

Kromě histonových modifikací přispívají k diverzitě a komplexitě histonového kódu také histonové varianty. Jedná se o samostatné histonové geny, které se od kanonické varianty liší některými aminokyselinami, což má vliv na jejich lokalizaci a funkci. U rostlin je nejlépe popsaným případem rodina variant histonu H2A, která zahrnuje H2A.Z, H2A.W a H2A.X (Amiard et al., 2010; Coleman-Derr and Zilberman, 2012b; Yelagandula et al., 2014). Kanonický histon H2A obvykle značí nukleosomy v rámci otevřeného a aktivního chromatinu a v exprimovaných genech. V 5' koncových oblastech stabilně exprimovaných genů a ve stresem či vývojově regulovaných genech pak bývá H2A nahrazován variantou H2A.Z (Coleman-Derr and Zilberman, 2012b; Yelagandula et al., 2014). Funkce H2A.Z není zcela jasná, ale předpokládá se, že reguluje transkripci a blokuje umístění DNA metylace. V heterochromatinových oblastech je dominantní variantou nedávno popsaná varianta H2A.W, která přispívá ke kondenzaci, opravám DNA a transkripční represi transpozonů (Yelagandula et al., 2014; Lorković et al., 2017; Osakabe et al., 2021). Poslední známou specializovanou variantou je H2A.X, jejíž fosforylovaná varianta značí místa poškození DNA (Amiard et al., 2010). Podobně jsou diverzifikovány také varianty histonů H3 a H1. Varianty H1.1 and H1.2 jsou u huseníčku rolního exprimovány během celého vývoje, zatímco H1.3 se exprimuje pouze ve specifických buňkách a je indukována vlivem stresových podmínek, kdy napomáhá regulaci genové exprese (Rutowicz et al., 2015). Rodina histon H3 genů zahrnuje u huseníčku rolního celkem 14 kopií, a to jednu kopii CENH3, pět kopií H3.1, tři kopie H3.3 a pět kopií H3.3-like (Okada et al., 2005). Centromerická histonová varianta CENH3 (známá také jako CENP-A) definuje funkční centromeru v rámci řady centromerických repetic a je charakteristická tím, že se vkládá do nukleosomu v průběhu G2 fáze buněčného cyklu (Lermontova et al., 2011; Ravi et al., 2011). Mutanti v CENH3 genu mají nerovnoměrnou segregaci chromosomů a sníženou fertilitu. Kanonický histon H3.1 je vkládán do nukleosomu histonovým chaperonem Chromatin Assembly Factor – 1 v průběhu DNA replikace. H3.1 je lokalizován po celém genomu, ale jeho

hustota je obecně vyšší v heterochromatiových oblastech (Stroud et al., 2012b). Mutanti s chybějícím funkčním komplexem Chromatin Assembly Factor – 1 mají pleiotropické fenotypy, včetně redukce heterochromatinu, částečné aktivace transpozonů a snížené stability genomu (Elmayan et al., 2005; Endo et al., 2006; Kirik et al., 2006; Mozgova and Hennig, 2015). Varianta H3.3 je naproti tomu exprimována během celého buněčného cyklu a nachází se především v transkripčně aktivních genech (Stroud et al., 2012b).

## 3.2 Molekulární analýzy vymezují čtyři základní typy chromatinu rostlin

Jak bylo popsáno výše, tak je v současnosti známo několik kontextů DNA metylace a celá řada histonových variant a modifikací. Bylo provedeno několik studií s cílem zhodnotit typy chromatinu na základě rozmístění a asociací různých modifikací a poskytnout určité diagnostické modifikace pro jednotlivé typy chromatinu. Analýzy u huseníčku rolního ukázaly, že lze rozlišit čtyři až devět hlavních typů chromatinu (Roudier et al., 2011; Sequeira-Mendes et al., 2014). Pro jednoduchost zde budou popsány pouze čtyři základní typy, které zahrnují: (i) aktivní geny, (ii) vývojově regulované geny, (iii) transpozony a (iv) mezigenové oblasti (**Tabulka 1**). Transkribované geny reprezentují typický euchromatin a obsahují množství transkripčně permisivních chromatinových modifikací: H3K4me2; H3K4me3; H3K9me3; H3K36me3; H3K56ac; H2Bub a CG DNA methylaci (**Obrázky 3C a 9**). Vývojově řízené geny nesou vysoký podíl transkripčně represivních modifikací H3K27me2 a H3K27me3 (**Obrázek 9**). Tyto geny obvykle alternují mezi reprimovanou a aktivovanou formou v průběhu vývoje rostliny.
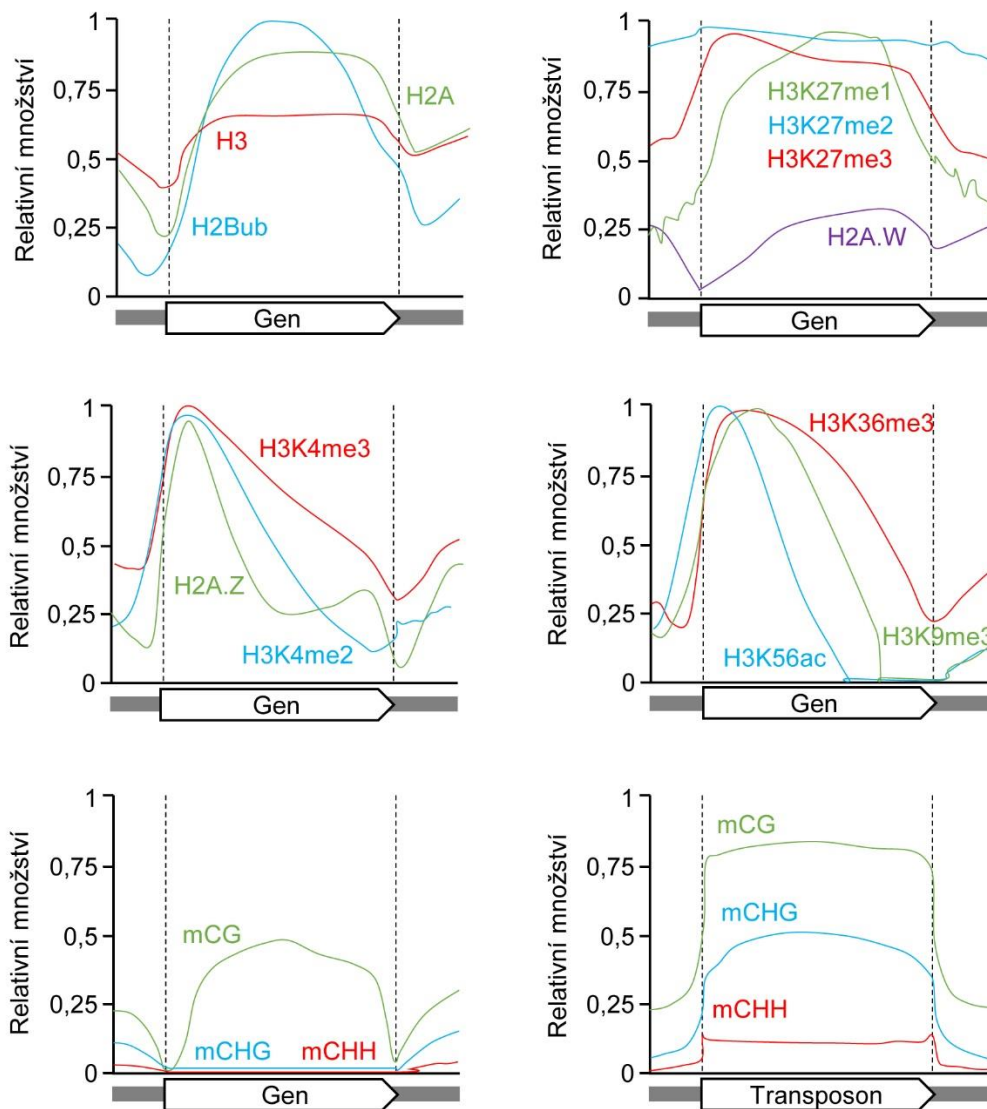
**Obrázek 9. Schématické rozmístění vybraných histonových variant a modifikací v genech.** X osa ukazuje gen, kde vertikální přerušované čáry značí počátek a konec transkribované oblasti. Y osa ukazuje relativní hustotu od 0 (chybí) do 1 (vysoce obohacen). Profily byly převzaty z publikací (Roudier et al., 2011; Coleman-Derr and Zilberman, 2012b; Yelagandula et al., 2014) s modifikacemi.

Oproti tomu transpozony, jiné repetitivní sekvence a malá část umlčených genů tvoří konstitutivní heterochromatin, který je reprimován stabilně pomocí transktipčního genového vypínaní a RdDM a je obohacen především o represivní modifikace včetně H3K9me2, H4K20me1 a DNA metylací ve všech sekvenčních kontextech. Poslední, poměrně nevyhraněný typ chromatinu pak tvoří mezigenové oblasti.

**Tabulka 1:** Základní typy chromatinu u huseníčku rolního (Roudier et al., 2011). Diagnostické modifikace jsou vyznačeny tučně.

| Typ | Charakteristické modifikace | Chromatin |
|---|---|---|
| Aktivní geny | **H2Bub**; **H3K4me2**; H3K4me3; **H3K9me3**; **H3K36me3**; H3K56ac; CG DNA methylace | Euchromatin |
| Vývojově kontrolované geny | **H3K27me3**; H3K27me2; H3K4me2 | Facultativní heterochromatin |
| Transpozony | **CHG and CHH DNA methylace**; **H2AW**; H3.1; **H3K9me2**; H3K27me1; H3K27me2; **H4K20me1** | Konstitutivní heterochromatin |
| Mezigenové oblasti | Absence výše popsaných. | Euchromatin |

Tato klasifikace je velmi užitečná s ohledem na diagnostiku nebo pochopení regulace genové exprese v různých oblastech genomu, za různých podmínek prostředí či v mutantních pozadích.

**4. VLIV ABIOTICKÉHO STRESU NA HETEROCHROMATIN ROSTLIN**

Jako přisedlé organismy musí rostliny reagovat na různé environentální podmínky bez možnosti úniku do příhodnějšího prostředí. Rostliny si proto vyvinuly řadu adaptací a strategií, které jim umožňují přežít i na velmi nehostinných místech a snášet extrémní abiotické podmínky včetně náhlých změn teploty či vysokých dávek UV záření. Klimatické modely navíc předpovídají, že zemědělství bude čelit stále extrémnějším výkyvům počasí, což může mít výrazné dopady na produkci potravin. Zatímco fyziologické a molekulární odpovědi rostlin vůči stresu jsou studovány v detailu, relativně málo je známo jak stress ovlivňuje rostlinný chromatin a epigenom. Toto téma se nicméně začíná dostávat do pořeldí zájmu a to jak u modelových druhů, tak i zemědělsky významných plodin (Gutzat and Mittelsten Scheid, 2012; **Pecinka and Mittelsten Scheid, 2012 - v této přehledné publikaci jsme shrnuli pohled na transgenerační epigenetické jevy a specifikovali pravidla pro jejich kritickou analýzu**; Bäurle, 2016; **Pecinka et al., 2020 - v této přehledné publikaci jsme shrnuli roli epigenomu ve vývoji a stresových odpovědích rostlin**).

## 4.1 Vliv stresu na strukturu a funkci heterochromatinu rostlin

Toto téma je v současnosti předmětěm intenzivního studia. Nejlépe jsou pak prostudovány vlivy teplotního a světelného stresu. Změny teploty mají výrazný dopad na rostlinný chromatin. Nízké teploty vedou především ke kompakci, zatímco vysoké způsobují spíše rozvolnění. Dlouhodobé vystavení semenáčů huseníčku rolního vysoké teplotě (37°C po dobu 24 až 30 h) způsobilo dekondenzaci heterochromatinových chromocenter v jádrech diferencovaných listových pletiv (**Pecinka et al., 2010 - v této práci jsme provedli detailní analýzu vlivu teplotního stresu na schopnost huseníčku kontrolovat transkripční represi transpozonů a heterochromatických chromocenter**). Tato dekondenzace byla stabilní a k obnovení původního stavu nedošlo ani po návratu do nestresových podmínek. Další neočekávaný fenotyp byla absence dekondenzace chromocenter v jádrech vrcholového apikálního meristému (**Pecinka et al., 2010**). To může souviset s faktem, že chromatinové proteiny jsou obecně více exprimovány v apikálních meristémech, což je pravděpodobně spojeno s vyšším stupněm organizace a stabilitou chromatinu v meristematických buňkách (Yadav et al., 2009; **Baubec et al., 2014 - v této studii jsme poukázali na funkci RdDM jako opravného mechanismu transkripční represe transpozonů a na zvýšnou expresi chromatinových genů v meristematických pletivech**). Na úrovni nukleosomů může vysoká teplota způsobit změny

v lokalizaci nebo četnosti, ale v současnosti není známo, zda je tento proces řízen enzymaticky nebo jen nižší vazebnou silou mezi histony a DNA v podmínkách vyšší teploty. Změny v chromatinové struktuře indukované vysokou teplotou mají vliv také na transkripci genů, nicméně toto pravidlo není univerzální (Kumar and Wigge, 2010; **Pecinka et al., 2010 - v rámci této publikace jsme dále ukázali, že zatímco u některých genů je transkripční aktivace doprovázena snížením nukleosomálního signálu, tak u jiných tomu tak není**). Zajímavým příkladem pozitivní regulace genové exprese pomocí nukleosomů je gen *HEAT SHOCK PROTEIN* (*HSP70*) u huseníčku rolního (Kumar and Wigge, 2010). Nukleosom pokrývající počátek transkripce *HSP70* obsahuje variantu H2A.Z a tak tlumí jeho transkripci. Zvyšování teploty pak koreluje se snižujícím se množstvím H2A.Z a rostoucím množstvím transkriptu. Z tohoto důvodu pojmenovali autoři teplotně senzitivní regulaci genové exprese u HSP70 jako „rostlinný termostat".

## 4.2 Může stres způsobit aktivaci a množení rostlinných transpozonů?

Přestože jsou transpozony reprimovány několika paralelními epigenetickými drahami, tak existují příklady recentních transpozičních událostí (Hu et al., 2011; Seymour et al., 2014; Willing et al., 2015). To naznačuje, že transpozony mohou alespoň občas uniknout epigenetické kontrole a množit se v hostitelském genomu. Námi objevený rozsah změn v organizaci heterochromatinu během teplotního stresu (**Pecinka et al., 2010 - v rámci této práce jsme ukázali dekondenzaci několika heterochromatických chromocenter vlivem teplotního stresu**) vedl k otázce, zda má tato změna vliv na efektivitu represe heterochromatinu. Myšlenka možného oslabení represivních vlastností repetitivní DNA během stresu je založena na pozorováních u několika druhů rostlin jako jsou např. hledík větší (*Antirrhinum majus*) nebo pomerančovník (*Citrus × sinensis*) (Hashida et al., 2015; Butelli et al., 2012). Gen *NIVEA*, který je u hledíku zodpovědný za barvu okvětí, obsahuje v promotorové oblasti retrotranspozon *Tam3*, který reguluje jeho expresi v závislosti na teplotě prostředí. Podobná situace pak panuje u tzv. „krvavých pomerančů", tj. odrůd pomerančovníků, jejichž dužina se barví červeně. Aby však došlo k červenému zbarvení dužiny, je nutné vystavit plody chladové periodě, během které dojde ke snížení represe specifického transpozonu v *cis*-regulatorní oblasti genu *RUBY*, který se aktivuje a způsobí akumulaci červeného pigmentu. U huseníčku byl testován vliv řady stresových faktorů na možnou aktivaci epigeneticky transkripčně vypnutého repetitivně organizovaného lokusu známého jako L5 nebo také TsGUS (Morel et al., 2000; Elmayan et al., 2005). V tomto případě však chlad, UV záření, osmotický stres, oxidativní stress ani indukce DNA zlomů bleomycinem nevedly k významnější aktivaci (A. Pecinka a O. Mittelsten Scheid,

nepublikovaná data). Jako nejúčinnější typ stresu, který vedl k poměrně silné aktivaci TsGUS lokusu se ukázal teplotní stres (24 až 30 h při 37°C) (**Pecinka et al., 2010**; Tittel-Elmer et al., 2010). TsGUS je uměle vloženou sekvencí, která nemusí dostatečně reprezentovat jiné repetitivní sekvence huseníčku, proto byla provedena celogenomová transkripční analýza za účelem zjistit zda teplotní stres aktivuje také endogenní repetitivní sekvence. Analýza pomocí expresních mikročipů odhalila, že asi 1% transpozonů huseníčku rolního vykazuje teplotně indukovanou transkripci. Podobná frekvence pak byla zjištěna také u příbuzného druhu *Arabidopsis lyrata* (**Pietzenuk et al., 2016 - v rámci této práce jsme analyzovali spektrum teplotně aktivovaných transpozonů u huseníčku písečného**), jehož genom obsahuje třikrát více repetic (Hu et al., 2011; **Rawat et al., 2015 - zde jsme provedli de novo anotaci genů a transpozonů v rámci genomu huseníčku písečného**). To naznačuje, že teplotní stres nezpůsobuje kompletní ztrátu epigenetické kontroly heterochromatinu, nicméně část transpozonů je teplotním stresem skutečně transkripčně aktivována. Většinu teplotně responsivních repetic tvořily LTR retrotranspozony a především rodina *COPIA78* (**Pecinka et al., 2010**; Tittel-Elmer et al., 2010; **Pietzenuk et al., 2016**). Tuto rodinu tvoří v referenčním genotypu *A. thaliana* Columbia celkem osm kopií, které jsou více než 1000-násobně transkripčně aktivovány teplotním stresem (**Pecinka et al., 2010**; Tittel-Elmer et al., 2010). *COPIA78* transkript byl detekovatelný dokonce ještě týden po skončení stresu, kdy byly ostatní teplotně indukované transpozony opět umlčeny. Na základě těchto vlastností byly *COPIA78* elementy pojmenovány jako *ONSEN*, což znamená japonsky „horké prameny" (Ito et al., 2011). Je zajímavé, že ztráta funkce udržovacích a *de novo* DNA metylačních drah nevedla k aktivaci transkripce z *ONSEN* elementů. Nicméně vystavení těchto mutantů teplotnímu stresu mělo aditivní efekt, který se projevoval více než 10000-násobnou aktivací. To naznačuje, že *ONSEN* elementy jsou řízeny či kontrolovány nejenom epigeneticky, ale i některou z teplotně responsivních regulačních drah. Toto zvláštní pozorování se následně podařilo vysvětlit zjištěním, že *LTR* úseky *ONSEN* elementů obsahují teplotně responsivní elementy (heat-responsive elements - HREs), na které se váže HEAT SHOCK FACTOR A 2 (HSFA2) (Cavrak et al., 2014). Funkční HRE se skládají ze shluku alespoň tří nGAAn sekvenčních motivů, které slouží jako vazebné místo nejméně tří molekul HSFA2. HSFA2 trimer pak tvoří komplex aktivního transkripčního faktoru. Výkon teplotně responsivního promotoru je určován jak počtem, tak i vzdáleností jednotlivých nGAAn motivů (Sakurai and Enoki, 2010):

- 4P HRE – obsahuje čtyři velmi blízko a pravidelně (2-4 bp) rozmístěné HRE motivy a podle navržených modelů může vázat až šest HSFA2 molekul. Promotory nesoucí 4P HRE vykazují nejvyšší stupeň aktivace po teplotním stresu.

- 3P HRE – obsahuje tři blízce a pravidelně rozmístěné motivy, váže tři HSFA2 molekuly.

- Gap HRE – obsahuje tři motivy, z nichž jeden je vzdálen až 7 bp od ostatních.
- Step HRE – obsahuje tři motivy, které jsou od sebe 7 bp a jsou jejich rozmístění je nepravidelné. Promotory obsahující tento typ motivu jsou nejslabší.

*ONSEN* retrotranspozony obsahují čtyři blízko a pravidelně rozmístěné nGAAn motivy (nGAAnnTTCnnnnGAAnnTTCn), což odpovídá 4P typu HRE. Přítomnost této klasické teplotně regulační sekvence je nejpravděpodobnějším důvodem, proč není *ONSEN/COPIA78* aktivován na pozadí epigenetických mutantů ve standardních, ale pouze v teplotně stresových podmínkách (**Pecinka et al., 2010**; Ito et al., 2011). Poněkud překvapivé bylo, že ani masivní navýšení transkriptu a dokonce přítomnost *ONSEN* cDNA nebyla dostatečná pro úspěšnou transpozici *ONSEN* elementů jak u standardních rostlin, tak u většiny RdDM mutantů (Ito et al., 2011). Jedinou vyjímku tvořily rostliny mutované v genu *NUCLEAR RNA POLYMERASE D 1* (*NRPD1*), který kóduje největší podjednotku DNA dependentní RNA polymerázy IV (PolIV), která zahajuje proces RdDM. Bez funkční PolIV nejsou *ONSEN* transkripty pravděpodobně rozeznávány jako transpozonové a nejsou tak zpracovány dalšími enzymy RdDM dráhy. Analýza potomstva *nrpd1* rostlin ukázala, že k integraci nových kopií dochází v apikálním meristému před diferenciací v samičí a samčí pohlavní orgány.

Absence nových kopií *ONSENu* v potomstvu teplotně stresovaných standardních rostlin vede k otázce, zda je zjištěná aktivace relevantní v přirozených podmínkách, a zda skutečně může vést k vyššímu počtu kopií daného transpozonu. Toto nelze v současnosti jednoznačně posoudit, protože provedené experimenty mají určitá omezení. Prvním je nízký počet (desítky) laboratorně testovaných rostlin, který nemohl podchytit méně častou (v řádu jednotek procent a nižší) frekvenci transpozice, která však může být potenciálně významná v podmínkách přirozené populace huseníčku čítající obrovské množství rostlin. Dále je pravděpodobné, že i v divoké populaci dochází vzácně k mutacím vedoucím ke ztrátě funkce některého z klíčových transkripčně represivních genů jako je *DDM1* or *NRPD1*. Přestože mutanti těchto genů nebudou v přírodních podmínkách dlouhodobě životaschopní, jejich příležitostný vznik (a pravděpodobně i relativně rychlý zánik) však může být u samosprašného druhu dostatečný pro namnožení transpozonů. Ty pak mohou občasným křížením mateřské rostliny s okolními rostlinami pronikat dále do populace. Za třetí, dlouhodobý teplotní stres aplikovaný v laboratoři neodpovídá přirozeným podmínkám, nicméně v přírodě lze očekávat kombinace teplotního stresu s dalšími faktory jako je sucho, zvýšené množství UV-B záření apod. Kombinace těchto faktorů na epigenetickou kontrolu transpozonů není známa, ale může mít pro rostlinu potenciálně oslabující efekt. Za čtvrté, srovnáním *LTR* úseků *ONSEN* elementů různých zástupců čeledi brukvovitých bylo zjištěno, že popsaný typ 4P HRE je starý nejméně 10 milionů let a vyskytuje se nejen u eurasijských, ale i severoamerických a australských zástupců čeledi

(**Pietzenuk et al., 2016**). To naznačuje, že přítomnost HRE může být pro *ONSEN* elementy evolučně výhodná. Na druhou stranu, existují také vyjímky, kdy u severoamerického druhu *Boechera stricta* došlo ke ztrátě funkčního 4P teplotně responzivního elementu a žádné *ONSEN*/*COPIA78* elementy nebyly zjištěny u eurasijského rodu *Capsella* (kokoška).

Naše recentní studie ukázala, že aktivace transpozonů teplotním stresem je poměrně častá mezi rodinami *COPIA* elementů. Kromě *COPIA78* (*ONSEN*) také u *COPIA37*, *TEMPERATURE RESPONSIVE TRANSPOSON* (*TERESTRA*) a *ROMANIAT5*. Fylogenetické analýzy dokazují, že tento znak vznikl opakovaně (**Pietzenuk et al., 2016**). Dále existují data naznačující, že podobný typ adaptace existuje i vůči jiným typům stresu. U vojtěšky seté (*Medicago sativa*) byl nalezen *MEDICAGO COLD-INDUCIBLE REPETITIVE ELEMENT* (*MCIRE*), jehož *LTR* obsahují chladově responsivní element (Ivashuta et al., 2002). Tento element má konzervovanou sekvenci CCGAC typickou pro C-repeat (CRT)/dehydračně responsivní elementy (DRE), které jsou rozeznávány chladově specifickými transkripčními faktory (Nakashima et al., 2009). To naznačuje, že uchování stresově specifických cis-regulačních elementů může být poměrně častou strategií, kterou uplatňují transpozony k přežití v rámci hostitelského genomu. Přítomnost stresově responsivních transpozonů by navíc mohla být evolučně výhodná také pro hostitelský genom. Při relativně nízké frekvenci transpozice (neohrožující stabilitu a funkce genomu) mohou stresově responsivní transpozony „rozsévat" své kopie, a s nimi i *cis*-regulační sekvence, které se mohou za určitých podmínek podílet na regulaci genů.

Zjištění, že některé transpozony využívají kanonické transkripční dráhy zodpovědné za reakce vůči stresu vede k provokativní otázce. Podařilo se transpozonům připojit také na regulační dráhy řídící vývoj rostlin? Napojení na vrcholové meristemy a reprodukční orgány by umožnilo množení přímo v pletivech vedoucích ke tvorbě příští generace bez nutnosti stresu. Tato hypotéza se zdá být podpořena nejméně dvěma pozorováními. Zaprvé, v *ddm1* mutantním pozadí je přibližně 1000 transpozonů transkripčně aktivováno bez indukovaného stresu (Zemach et al., 2013). Jejich aktivace tedy není vázána na stresové transkripční faktory (Miura et al., 2001; Mirouze et al., 2009; Tsukahara et al., 2009). Zadruhé, analýza vazebných míst květního transkripčního faktoru SEPALLATA3 pomocí metody ChIP-seq odhalila, že se kromě standardních genů váže také na řadu transpozonů obsahující jeho typické vazebné motivy (Muiño et al., 2011). Tato pozorování by v konečném důsledku mohla znamenat, že transpozony vážící stresově specifické transkripční faktory využívají méně efektivní strategii než ty parazitující na vývojových drahách.

Lze tedy úspěch transpozonových rodin aktivovaných stresem měřit počtem jejich kopií? Na tuto otázku není snadné zodpovědět, z důvodu několika obtížně měřitelných faktorů

jako např. dynamika eliminace transpozonů. V situaci, kdy bude úroveň transpozice i eliminate poměrně nízká, může transpozonová rodina vykazovat poměrně stabilní počet kopií. Naopak, pokud bude docházet k rychlé amplifikaci, ale ještě rychlejší eliminaci, se bude počet kopií snižovat. Zde může posloužit jako vodítko podobnost LTR sekvencí na obou koncích LTR retrotranspozonů, protože tyto sekvence jsou ve chvíli inzerce identické a časem nezávisle akumulují mutace. Všechny dosud známé teplotně responzivní transpozonové rodiny mají poměrně malý počet kopií (<100 kompletních kopií/genom) a na základě jejich LTR sekvencí se jeví jako evolučně mladé (>90% identita) (Cavrak et al., 2014; **Pietzenuk et al., 2016**). To naznačuje, že teplotně responsivní rodiny transpozonů mají relativně rychlou obměnu kopií v genomu.

V souhrnu, recentní výzkum odpovědí transpozonů na stres poskytl řadu nových poznatků o vztazích mezi těmito genomickými paratzity a jejich hostitelskými genomy. Velkým překvapením bylo, jak je tato mapa vztahů barvitá a plastická. Současně v této oblasti zůstává řada neznámých, které jsou předmětem aktivního výzkumu.

# 5. JAKÁ JE BUDOUCNOST STUDIA CHROMATINU ROSTLIN?

Závěrem své práce si dovolím nastínit možné směry výzkumu rostlinného (hetero)chromatinu a regulace genové exprese. V současnosti, 93 let po publikaci objevu euchromatinu a heterochromatinu německým biologem Emilem Heitzem (Heitz, 1928), známe základní biochemickou povahu chromatinu, poměrně velké množství proteinů chromatinu a také řadu drah, které jsou zodpovědné za instalaci či odstraňování určitých post translačních modifikací. Zvláště během posledních dvaceti let došlo k úžasnému pokroku. To bylo umožněno kombinací řady příznivých faktorů včetně vytvoření referenčních sekvencí genomů hlavních modelových rostlin, veřejných sbírek mutantů a mnoha různých technik analýzy genové exprese a detekce chromatinových modifikací. Lze říci, že se momentálně nacházíme ve zlaté éře studia chromatinu, kdy jsme pochopili základní principy řady epigenetických drah, nicméně detailní poznání a následně aplikace výsledků do biotechnologické či zemědělské praxe je stále před námi.

Mezi stálice epigenetického výzkumu patří identifikace nových chromatinových značek a následně jejich funkční analýza. I v současnosti se počet známých modifikací stále rozšiřuje a lze předpokládat, že toto zůstane jedním z pilířů tohoto oboru. Mezi nové kandidáty patří např. metylace N6-adeninu, specifické histonové varianty z okruhu histonu H2A nebo H1. Dosud málo prozkoumány jsou konformační modifikace DNA - A-, B- a Z-varianty, G-kvadruplex nebo i-motivy. Další téměř nedotčenou skupinou jsou modifikace RNA, které budou hrát velmi důležitou roli v řadě epigenetických procesů. Detailní studium těchto modifikací bude vyžadovat kombinaci existujících přístupů (genetické skríny, proteomické a biochemické analýzy, vývoj protilátek, celogenomové analýzy apod.), jakož i vývoj nových metod.

O stupeň výše bude nutné propojovat znalosti o rozmístění, dynamice a funkcích jednotlivých značek do charakteristických chromatinových stavů. Na základě existujících studií lze předpokládat, že tato část výzkumu bude poměrně složitá. Rozsáhlé studie často kombinují materiály a chromatinové profily generované v různých experimentálních podmínkách (včetně kontrolních), na různých pletivech a někdy i genotypech. To vede nutně k větší heterogenitě mezi vzorky a může se projevit na nižší citlivosti experimentů. U rostlin se k tomuto tradičně přidává heterogenita pletiv, takže výsledný signál je pak jakýmsi průměrem daného vzorku.

Používání mixu buněk z různých pletiv je dáno strukturou rostlinných orgánů, nemožností kultivace jednotlivých buněčných typů jako u živočichů a také množstvím materiálu nutného pro řadu epigenetických experimentů (obvykle stovky miligramů až gramy). Tento problém je nyní alespoň částečně eliminován novými protokoly pro analýzu jednotlivých buněk (single cell). Příklady jsou studie komplexních pletiv jako je kořen či endosperm, kde

analýza založená na sekvenování molekul z jednotlivých buněk či jader odhalila nové expresní programy či skupiny buněk s dosud neznámými funkcemi (Dorrity et al., 2021; Picard et al., 2021).

Co bylo dříve, slepice nebo vejce? Varianty této otázky si klade řada epigenetiků a určení posloupnosti dějů zůstává jedním z klasických problémů epigenetických studií. Přítomnost řady chromatinových modifikací je silně korelována, což znesnadňuje identifikaci těch, které tvoří základ určitých chromatinových stavů. Indukuje navázání transkripčního faktoru změnu z heterochromatinu na euchromatin nebo musí nejdříve dojít ke změně a teprve poté může dojít k navázání transkripčního faktoru? Postupným detailním studiem jednotlivých drah se však daří osvětlovat i tyto problémy. Fascinujícím příkladem zůstává dosud probíhající diskuse zda je transkripční vypínání genů indukováno dříve DNA nebo histony, resp. jejich modifikacemi (Osakabe et al., 2021).

Většina našich znalostí týkajících se organizace, povahy a funkce rostlinného chromatinu je založena na studiích provedených u huseníčku rolního. Huseníček je výborný modelový systém, který má rozsáhlé genetické, molekulární a genomické zdroje velmi vysoké kvality. Na druhou stranu je genom huseníčku rolního do jisté míry atypický svou malou velikostí (okolo 150 Mbp; referenční sekvence má 119 Mbp, ale neobsahuje některé tandemově repetitivních oblastí, především centromery a ribozomální DNA), nízkým obsahem repetitivních sekvencí a malým množstvím heterochromatinu nahloučeným do pericentromerických oblastí. Jedním z perspektivních směrů rostlinné epigenetiky je jistě analýza chromatinu u jiných druhů. Dlouhodobě jsou známy rozdíly v globální organizaci buněčného jádra. Zde jsou protipólem huseníčku druhy s tzv. Rabl organizací. Molekulární metody však začínají odhalovat zcela novou vrstvu variability. Příkladem může být analýza genové metylace u různých druhů čeledi brukvovitých (Bewick and Schmitz, 2017). Zatímco u huseníčku rolního nese CG DNA metylaci přibližně 20% genů, tak tato metylace zcela chybí u některých jiných druhů a zdá se, že koreluje s absencí funkční CMT3. To je velmi zvláštní pozorování, protože CMT3 zajišťuje u huseníčku rolního CHG metylaci a nikoliv CG metylaci. Lze předpokládat, že takovýchto rozdílů existuje celá řada, ale dosud nejsou známy. Dále víme poměrně málo, kdy se epigenetické mechanismy podílí na regulaci hospodářsky významných znaků. Mezi dosud známé příklady patří epigenetická kontrola pohlaví melounů, vývoje plodu palmy olejnaté, pimentace dužiny „krvavých pomerančů" nebo kontrola vernalizace u obilovin (Martin et al., 2009; Oliver et al., 2009; Butelli et al., 2012; Ong-abdullah et al., 2015). Tyto znaky jistě představují pouhou špičku ledovce epigenetické variability a regulace přítomné v plodinách a do budoucna by je bylo možné použít pro potřeby šlechtění.

Trvalou výzvou zůstává analýza chromatinu u druhů s velkým genomem jako je kukuřice (2.5 Gb) nebo ječmen (5.1 Gb) a ještě o řád obtížnější je pak u polyploidů jako je pšenice setá (17 Gb). I zde však dochází k vývoji dostupných zdrojů (Concia et al., 2020), což je umožněno především vyšší účinností sekvenovacích technik nové generace a stále se zlepšujícími programy pro zpracování dat. Nicméně celogenomová analýza rozmístění určitých chromatinových modifikací u druhů s velkým genomem je stále velmi drahou záležitostí. Pro srovnání, 40-ti násobné sekvenování metylomu huseníčku lze pořídit za cenu okolo 5000 Kč. V případě ječmene se jedná již o přibližně 200 000 Kč a u pšenice o 650 000 Kč. To značí, že řadu experimentů je nutno zacílit na určitou část genomu a nadále vyvíjet nové levnější metody sekvenování DNA.

Jeden z přehlížených aspektů rostlinné chromatinové biologie je, jak chromatin přispívá se stabilitě genomu. Stále větší množství studií naznačuje, že určité typy chromatinu mohou být více či méně náchylné k akumulaci poškození DNA a mutací. Heterochromatin obecně vykazuje nižší míru oprav DNA a ztráta některých faktorů upravujících strukturu a funkci heterochromatinu vede ke zvýšené citlivosti k DNA poškozujícím látkám (Shaked et al., 2006; Jacob et al., 2009; Rosa and Shaw, 2013; Donà and Scheid, 2015; **Liu et al., 2015 - v této práci jsme ukázali, že cytidinový analog zebularin indukuje nejen DNA demetylaci, ale také dosud neznámé poškození DNA**; **Willing et al., 2016 - zde jsme sekvenovali genomy několika genotypů huseníčku vystavených dlouhodobému působení simulovaného slunečního záření a tím jsme odhalili predominantní typy mutací vznikající v podmínkách standardních a fotoreparačně defektních rostlin**). Lokalizace DNA zlomů pro homologní rekombinaci se zdá být také určena epigeneticky (Choi et al., 2013). Interakce chromatinových a DNA reparačních superdrah je proto jedním z budoucích velmi zajímavých témat kterými se bude výzkum buněčného jádra také ubírata této oblasti se aktivně věnuje také moje výzkumná skupina. Především s ohledem na roli Strukturního komplexu údržby chromosomů 5/6 (Structural maintenance of chromosomes 5/6; SMC5/6) v opravách určitých poškození DNA a transgenerační stabilitě genomu rostlin (**Liu et al., 2015; Diaz et al., 2019 - v této práci jsme funkčně charakterizovali NSE4 jako podjednotku SMC5/6 komplexu u huseníčku rolního**; Yang et al., 2021).

Kritická analýza a interpretace výsledků je zásadní v jakémkoliv vědním oboru a epigenetický výzkum není vyjímkou. Někteří výzkumníci mají tendence vnímat chromatin jako téměř magickou substanci, která má (v nadsázce) schopnost pamatovat si minulost a vidět do budoucnosti (**Pecinka and Mittelsten Scheid, 2012 - v tomto přehledném článku jsme kriticky posuzovali možnosti epigenetické mezigenerační paměti**). Přestože některé epigenetické procesy vykazují určitou stochasticitu, obecně se jedná o přesně regulované

biochemické reakce, které probíhají s vysokou předvídatelností. Míra toho co se v určité chvíli jeví náhodným pak dále klesá s postupným odkrýváním molekulárních mechanismů epigenetických procesů.

Na základě tohoto výčtu jsem přesvědčen, že studium buněčného jádra, heterochromatinu a regulace genové exprese má vysoký potenciál generovat nové zásadní poznatky u modelových i hospodářsky významných druhů rostlin. Postupem času lze také očekávat čím dále větší aplikaci výsledků studia chromatinu pro rostlinné biotechnologie a šlechtění rostlin.

# 6. LITERATURA

**Abdelsamad, A. and Pecinka, A.** (2014). Pollen-specific activation of Arabidopsis retrogenes is associated with global transcriptional reprogramming. Plant Cell **26**: 3299–3313.

**Alabert, C. and Groth, A.** (2012). Chromatin replication and epigenome maintenance. Nat. Rev. Mol. Cell Biol. **13**: 153–167.

**Alberts, B.** (2002). Chromosomal DNA and its packaging in the chromatin fiber. In Molecular Biology of the Cell, B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, eds (Garland science: New York).

**Ambrožová, K., Mandáková, T., Bureš, P., Neumann, P., Leitch, I.J., Koblížková, A., Macas, J., and Lysak, M.A.** (2011). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of Fritillaria lilies. Ann. Bot. **107**: 255–268.

**Amiard, S., Charbonnel, C., Allain, E., Depeiges, A., White, C.I., and Gallego, M.E.** (2010). Distinct roles of the ATR kinase and the Mre11-Rad50-Nbs1 complex in the maintenance of chromosomal stability in Arabidopsis. Plant Cell **22**: 3020–3033.

**Appels, R. et al.** (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science (80-. ). **361**: eaar7191.

**Arabidopsis Genome Initiative** (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408**: 796–815.

**Arkhipova, I.R.** (2006). Distribution and phylogeny of Penelope-like elements in eukaryotes. Syst. Biol **55**: 875–885.

**Arteaga-Vazquez, M.A. and Chandler, V.L.** (2010). Paramutation in maize: RNA mediated trans-generational gene silencing. Curr. Opin. Genet. Dev. **20**: 156–163.

**Bartee, L., Malagnac, F., Bender, J., Bartee, L., Malagnac, F., and Bender, J.** (2001). Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev. **15**: 1753–1758.

**Bártová, E., Krej, J., Harni, A., Galiová, G., and Kozubek, S.** (2008). Histone Modifications and Nuclear Architecture : A Review. **56**: 711–721.

**Batista, R.A. and Köhler, C.** (2020). Genomic imprinting in plants—revisiting existing models. Genes Dev. **34**: 24–36.

**Baubec, T., Dinh, H.Q., Pecinka, A., Rakic, B., Rozhon, W., Wohlrab, B., von Haeseler, A., and Mittelsten Scheid, O.** (2010). Cooperation of multiple chromatin modifications can generate unanticipated stability of epigenetic states in Arabidopsis. Plant Cell **22**: 34–47.

**Baubec, T., Finke, A., Mittelsten Scheid, O., and Pecinka, A.** (2014). Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. EMBO Rep. **15**: 446–452.

**Bäurle, I.** (2016). Plant Heat Adaptation: Priming in response to heat stress. F1000Research **5**: 1–5.

**Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A.** (2009). An operational definition of epigenetics. Genes Dev **23**: 781–783.

**Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E.** (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One **3**.

**Berr, A., Pecinka, A., Meister, A., Kreth, G., Fuchs, J., Blattner, F.R., Lysak, M.A., and**

**Schubert, I.** (2006). Chromosome arrangement and nuclear architecture but not centromeric sequences are conserved between Arabidopsis thaliana and Arabidopsis lyrata. Plant J. **48**: 771–783.

**Bewick, A.J., Niederhuth, C.E., Ji, L., Rohr, N.A., Griffin, P.T., Leebens-Mack, J., and Schmitz, R.J.** (2017). The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol. **18**: 65.

**Bewick, A.J. and Schmitz, R.J.** (2017). Gene body DNA methylation in plants. Curr. Opin. Plant Biol. **36**: 103–110.

**Van Blokland, R., Van der Geest, N., Mol, J.N.M., and Kooter, J.M.** (1994). Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover. Plant J. **6**: 861–877.

**Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-recupero, G., and Martin, C.** (2012). Retrotransposons Control Fruit-Specific , Cold-Dependent Accumulation of Anthocyanins in Blood Oranges.

**Casacuberta, J.M. and Santiago, N.** (2003). Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene **311**: 1–11.

**Cavrak, V. V, Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M., and Mittelsten Scheid, O.** (2014). How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. PLOS Genet. **10**: e1004115.

**Chan, S.W.L., Henderson, I.R., and Jacobsen, S.E.** (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat. Rev. Genet. **6**: 351–360.

**Chandler, V. and Alleman, M.** (2008). Anecdotal , Historical and Critical Commentaries on Genetics. **1844**: 1839–1844.

**Chandler, V.L. and Stam, M.** (2004). Chromatin conversations: Mechanisms and implications of paramutation. Nat. Rev. Genet. **5**: 532–544.

**Chanou, A. and Hamperl, S.** (2021). Single-Molecule Techniques to Study Chromatin. Front. Cell Dev. Biol. **9**: 1596.

**Chen, H., He, C., Wang, C., Wang, X., Ruan, F., Yan, J., Yin, P., Wang, Y., and Yan, S.** (2021). RAD51 supports DMC1 by inhibiting the SMC5/6 complex during meiosis. Plant Cell.

**Chiolo, I., Minoda, A., Colmenares, S.U., Polyzos, A., Costes, S. V, and Karpen, G.H.** (2011). Double-Strand Breaks in Heterochromatin Move Outside of a Dynamic HP1a Domain to Complete Recombinational Repair. Cell **144**: 732–744.

**Chodavarapu, R.K. et al.** (2010). Relationship between nucleosome positioning and DNA methylation. Nature **466**: 388–392.

**Choi, K., Zhao, X., Kelly, K.A., Venn, O., Higgins, J.D., Yelina, N.E., Hardcastle, T.J., Ziolkowski, P.A., Copenhaver, G.P., Franklin, F.C.H., Mcvean, G., and Henderson, I.R.** (2013). Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat. Genet. **45**: 1327–1338.

**Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L.** (2002). DEMETER , a DNA Glycosylase Domain Protein , Is Required for Endosperm Gene Imprinting and Seed Viability in Arabidopsis. **110**: 33–42.

**Coleman-Derr, D. and Zilberman, D.** (2012a). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. PLoS Genet **8**: e1002988.

**Coleman-Derr, D. and Zilberman, D.** (2012b). Deposition of Histone Variant H2A.Z within Gene Bodies Regulates Responsive Genes. PLoS Genet. **8**.

**Comai, L.** (2005). The advantages and disadvantages of being polyploid. Nat. Rev. Genet. **6**: 836–846.

**Concia, L. et al.** (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol. **21**: 104.

**Copenhaver, G.P. and Pikaard, C.S.** (1996). RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of Arabidopsis thaliana adjoin the telomeres on chromosomes 2 and 4. Plant J **9**: 259–272.

**Craig, N.L., Green, R., Greider, C., Storz, G., Wolberger, C., and Orna, C.-F.** (2014). Molecular biology: Principles of genome function 2nd ed. (Glasgow).

**Daxinger, L., Hunter, B., Sheikh, M., Jauvion, V., Gasciolli, V., Vaucheret, H., Matzke, M., and Furner, I.** (2008). Unexpected silencing effects from T-DNA tags in Arabidopsis. Trends Plant Sci. **13**: 4–6.

**Diaz, M., Pecinkova, P., Nowicka, A., Baroux, C., Sakamoto, T., Yuliani Gandha, P., Jeřábková, H., Matsunaga, S., Grossniklaus, U., and Pecinka, A.** (2019). SMC5/6 complex subunit NSE4A is involved in DNA damage repair and seed development in Arabidopsis. Plant Cell **31**: 1579–1597.

**Doğan, E.S. and Liu, C.** (2018). Three-dimensional chromatin packing and positioning of plant genomes. Nat. Plants **4**: 521–529.

**Donà, M. and Scheid, O.M.** (2015). Update on DNA Damage Repair DNA Damage Repair in the Context of Plant Chromatin 1. **168**: 1206–1218.

**Dorrity, M.W., Alexandre, C.M., Hamm, M.O., Vigil, A.-L., Fields, S., Queitsch, C., and Cuperus, J.T.** (2021). The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nat. Commun. **12**: 3334.

**van Driel, R., Fransz, P.F., and Verschure, P.J.** (2003). The eukaryotic genome: a system regulated at different hierarchical levels. J Cell Sci **116**: 4067–4075.

**Du, J. et al.** (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151**: 167–180.

**Elmayan, T., Proux, F., and Vaucheret, H.** (2005). Arabidopsis RPA2: A genetic link among transcriptional gene silencing, DNA repair, and DNA replication. Curr. Biol. **15**: 1919–1925.

**Endo, M., Ishikawa, Y., Osakabe, K., Nakayama, S., Kaya, H., Araki, T., Shibahara, K.I., Abe, K., Ichikawa, H., Valentine, L., Hohn, B., and Toki, S.** (2006). Increased frequency of homologous recombination and T-DNA integration in Arabidopsis CAF-1 mutants. EMBO J. **25**: 5579–5590.

**Ernst, J. et al.** (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature **473**: 43–49.

**Fajkus, P. et al.** (2019). Telomerase RNAs in land plants. Nucleic Acids Res. **47**: 9842–9856.

**Fajkus, P., Peška, V., Sitová, Z., Fulnečková, J., Dvořáčková, M., Gogela, R., Sýkorová, E., Hapala, J., and Fajkus, J.** (2016). Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)n is synthesized by telomerase. Plant J **85**: 337–347.

**Fawcett, J.A., Maere, S., and Van De Peer, Y.** (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc. Natl. Acad. Sci. U. S. A. **106**: 5737–5742.

**Feng, S., Jacobsen, S.E., and Reik, W.** (2010). Epigenetic reprogramming in plant and animal development. Science **330**: 622–627.

**Feschotte, C., Jiang, N., and Wessler, S.R.** (2002). Plant transposable elements: Where genetics meets genomics. Nat. Rev. Genet. **3**: 329–341.

**Finke, A., Mandáková, T., Nawaz, K., Vu, G.T.H., Novák, P., Macas, J., Lysak, M.A., and Pecinka, A.** (2019). Genome invasion by a hypomethylated satellite repeat in Australian crucifer Ballantinia antipoda. Plant J. **99**: 1066–1079.

**Finnegan, E.J., Peacock, W.J., and Dennis, E.S.** (1996). Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. Proc. Natl. Acad. Sci. U. S. A. **93**: 8449–8454.

**Fu, Y. et al.** (2015). N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. Cell **161**: 879–892.

**Fuchs, J., Demidov, D., Houben, A., and Schubert, I.** (2006). Chromosomal histone modification patterns - from conservation to diversity. Trends Plant Sci. **11**: 199–208.

**Fultz, D., Choudury, S.G., and Slotkin, R.K.** (2015). Silencing of active transposable elements in plants. Curr. Opin. Plant Biol. **27**: 67–76.

**Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldán-Arjona, T., David, L., and Zhu, J.K.** (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell **111**: 803–814.

**Gregory, T.R.** (2005). The C-value Enigma in Plants and Animals : A Review of Parallels and an Appeal for Partnership.: 133–146.

**Groth, M. et al.** (2016). MTHFD1 controls DNA methylation in Arabidopsis. Nat. Commun. **7**.

**Gutzat, R. and Mittelsten Scheid, O.** (2012). Epigenetic responses to stress: Triple defense? Curr. Opin. Plant Biol. **15**: 568–573.

**Habermann, F.A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., and Solovei, I.** (2001). Arrangements of macro- and microchromosomes in chicken cells. Chromosom. Res **9**: 569–584.

**Hashida, S., Kitamura, K., Mikami, T., and Kishima, Y.** (2015). Temperature Shift Coordinately Changes the Activity and the Methylation State of Transposon Tam3 in Antirrhinum majus.: 1207–1216.

**Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., and Cheng, X.** (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature **455**: 826–829.

**Havecker, E.R., Gao, X., and Voytas, D.F.** (2004). The diversity of LTR retrotransposons. Genome Biol **5**: 1–6.

**Heitz, E.** (1928). Das Heterochromatin der Moose. Jahrbücher für wissenschaftliche Bot. **69**: 762 – 818.

**Hollister, J.D., Smith, L.M., Guo, Y.L., Ott, F., Weigel, D., and Gaut, B.S.** (2011). Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proc. Natl. Acad. Sci. U. S. A. **108**: 2322–2327.

**Hsieh, T., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D.** (2009). Genome-Wide Demethylation of Arabidopsis Endosperm. Science (80-. ). **324**: 1451–1454.

**Hu, T.T. et al.** (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat. Genet. **43**: 476–483.

**Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A.J.M., and Matzke, M.** (2006). Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. EMBO J. **25**: 2828–2836.

**Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J.** (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature **472**: 115–120.

**Ivashuta, S., Naumkina, M., Gau, M., Uchiyama, K., Isobe, S., Mizukami, Y., and Shimamoto, Y.** (2002). Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (Medicago sativa). Plant J. cell Mol. Biol. **31**: 615–627.

**Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y. V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D.** (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat. Struct. Mol. Biol. **16**: 763–768.

**Jasencakova, Z., Meister, A., and Schubert, I.** (2001). Chromatin organization and its relation to replication and histone acetylation during the cell cycle in barley. Chromosoma **110**: 83–92.

**Jeddeloh, J.A., Stokes, T.L., and Richards, E.J.** (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat. Genet. **22**: 94–97.

**Jeppsson, K., Kanno, T., Shirahige, K., and Sjögren, C.** (2014). The maintenance of chromosome structure: Positioning and functioning of SMC complexes. Nat. Rev. Mol. Cell Biol. **15**: 601–614.

**Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J., Hodnett, G., Drabek, J., Lopez, R., and Price, H.J.** (2005). Evolution of genome size in Brassicaceae. Ann Bot **95**: 229–235.

**Kaessmann, H., Vinckenbosch, N., and Long, M.** (2009). RNA-based gene duplication: Mechanistic and evolutionary insights. Nat. Rev. Genet. **10**: 19–31.

**Kawakatsu, T. et al.** (2016). Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell **166**: 492–505.

**Kazda, A., Zellinger, B., Rössler, M., Derboven, E., Kusenda, B., and Riha, K.** (2012). Chromosome end protection by blunt-ended telomeres. Genes Dev **26**: 1703–1713.

**Kellogg, E.A. and Bennetzen, J.L.** (2004). The evolution of nuclear genome structure in seed plants. Am J Bot **91**: 1709–1725.

**Khandelwal, S.** (1990). Chromosome evolution in the genus Ophioglossum L. Bot J Linn Soc **102**: 205–217.

**Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., and Kakutani, T.** (2007). Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. Plant J **49**: 38–45.

**Kirik, A., Pecinka, A., Wendeler, E., and Reiss, B.** (2006). The chromatin assembly factor subunit FASCIATA1 is involved in homologous recombination in plants. Plant Cell **18**: 2431–2442.

**Köhler, C., Wolff, P., and Spillane, C.** (2012). Epigenetic mechanisms underlying genomic imprinting in plants. Annu. Rev. Plant Biol. **63**: 331–352.

**Kouzarides, T.** (2007). Chromatin Modifications and Their Function. Cell **128**: 693–705.

**Kumar, S.V. and Wigge, P.A.** (2010). H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. Cell **140**: 136–147.

**Law, J.A. and Jacobsen, S.E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. **11**: 204–220.

**Leach, C.R., Donald, T.M., Franks, T.K., Spiniello, S.S., Hanrahan, C.F., and Timmis, J.N.** (1995). Organisation and origin of a B chromosome centromeric sequence from Brachycome dichromosomatica. Chromosoma **103**: 708–714.

**Lee, J., Jang, H., Shin, H., Choi, W.L., Mok, Y.G., and Huh, J.H.** (2014). AP endonucleases process 5-methylcytosine excision intermediates during active DNA demethylation in Arabidopsis. Nucleic Acids Res. **42**: 11408–11418.

**Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., Koszegi, D., and Schubert, I.** (2011). Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. Plant J. **68**: 40–50.

**Lermontova, I., Sandmann, M., and Demidov, D.** (2014). Centromeres and kinetochores of Brassicaceae.: 135–152.

**Li, B., Carey, M., and Workman, J.L.** (2007). The Role of Chromatin during Transcription. Cell **128**: 707–719.

**Li, Y., Duan, C.-G., Zhu, X., Qian, W., and Zhu, J.-K.** (2015a). A DNA ligase required for active DNA demethylation and genomic imprinting in Arabidopsis. Cell Res. **25**: 757–760.

**Li, Z., Baniaga, A.E., Sessa, E.B., Scascitelli, M., Graham, S.W., Rieseberg, L.H., and Barker, M.S.** (2015b). Early genome duplications in conifers and other seed plants. Sci. Adv. **1**: 1–8.

**Lisch, D.** (2013). How important are transposons for plant evolution? Nat. Rev. Genet. **14**: 49–61.

**Liu, C., Lu, F., Cui, X., and Cao, X.** (2010). Histone Methylation in Higher Plants. Annu. Rev. Plant Biol. **61**: 395–420.

**Liu, C.H., Finke, A., Díaz, M., Rozhon, W., Poppenberger, B., Baubec, T., and Pecinka, A.** (2015). Repair of DNA damage induced by the cytidine analog zebularine requires ATR and ATM in arabidopsis. Plant Cell **27**: 1788–1800.

**Long, Q. et al.** (2013). Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat. Genet. **45**: 884–890.

**Lorković, Z.J., Park, C., Goiser, M., Jiang, D., Kurzbauer, M.T., Schlögelhofer, P., and Berger, F.** (2017). Compartmentalization of DNA Damage Response between Heterochromatin and Euchromatin Is Mediated by Distinct H2A Histone Variants. Curr. Biol. **27**: 1192–1199.

**Luo, G.Z., Blanco, M.A., Greer, E.L., He, C., and Shi, Y.** (2015). DNA N6-methyladenine: A new epigenetic mark in eukaryotes? Nat. Rev. Mol. Cell Biol. **16**: 705–710.

**Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I.** (2006). Mechanisms of chromosome number reduction in Arabidopsis thaliana and related Brassicaceae species. Proc. Natl. Acad. Sci. U. S. A. **103**: 5224–5229.

**Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K., and Lysaka, M.A.** (2010). Fast diploidization in close mesopolyploid relatives of Arabidopsis. Plant Cell **22**: 2277–2290.

**Marí-ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O.** (2013). Reconstructing de novo silencing of an active plant retrotransposon.

**Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., Pitrat, M., Dogimont, C., and Bendahmane, A.** (2009). A transposon-induced epigenetic change leads to sex determination in melon. Nature **461**: 1135–1138.

**Martínez-Macías, M.I., Qian, W., Miki, D., Pontes, O., Liu, Y., Tang, K., Liu, R., Morales-Ruiz, T., Ariza, R.R., Roldán-Arjona, T., and Zhu, J.-K.** (2012). A DNA 3′ Phosphatase Functions in Active DNA Demethylation in Arabidopsis. Mol. Cell **45**: 357–370.

**Mascher, M. et al.** (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature **544**: 427–433.

**Mathieu, O., Reinders, J., Čaikovski, M., Smathajitt, C., and Paszkowski, J.** (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell **130**: 851–862.

**Matzke, M.A. and Mosher, R.A.** (2014). RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. Nat. Rev. Genet. **15**: 394–408.

**McClintock, B.** (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U.S.A. **36**: 344–355.

**Melamed-Bessudo, C., Yehuda, E., Stuitje, A.R., and Levy, A.A.** (2005). A new seed-based assay for meiotic recombination in Arabidopsis thaliana. Plant J. **43**: 458–466.

**Melters, D.P. et al.** (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. **14**: R10.

**Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J., and Mathieu, O.** (2009). Selective epigenetic control of retrotransposition in Arabidopsis. Nature **461**: 427–430.

**Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T.** (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature **411**: 212–214.

**Morel, J.B., Mourrain, P., Béclin, C., and Vaucheret, H.** (2000). DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. Curr. Biol. **10**: 1591–1594.

**Mozgova, I. and Hennig, L.** (2015). The Polycomb Group Protein Regulatory Network. Annu. Rev. Plant Biol. **66**: 269–296.

**Muiño, J.M., Kaufmann, K., van Ham, R.C.H.J., Angenent, G.C., and Krajewski, P.** (2011). ChIP-seq Analysis in R (CSAR): An R package for the statistical detection of protein-bound genomic regions. Plant Methods **7**: 11.

**Nabel, C.S., Manning, S.A., and Kohli, R.M.** (2012). The curious chemical biology of cytosine: Deamination, methylation, and oxidation as modulators of genomic potential. ACS Chem Biol **7**: 20–30.

**Nakashima, K., Ito, Y., and Yamaguchi-Shinozaki, K.** (2009). Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol. **149**: 88–95.

**Okada, T., Endo, M., Singh, M.B., and Bhalla, P.L.** (2005). Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Plant J. **44**: 557–568.

**Oliver, S.N., Finnegan, E.J., Dennis, E.S., Peacock, W.J., and Trevaskis, B.** (2009). Vernalization-induced flowering in cereals is associated with changes in histone methylation at the VERNALIZATION1 gene. Proc. Natl. Acad. Sci. U. S. A. **106**: 8386–8391.

**Ong-abdullah, M. et al.** (2015). the mantled somaclonal variant of oil palm.

**Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S.A., Akimcheva, S., Kuehn, A.L.,**

**Pisupati, R., Lorković, Z.J., Yelagandula, R., Kakutani, T., and Berger, F.** (2021). The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. Nat. Cell Biol. **23**: 391–400.

**Paterson, A.H., Freeling, M., and Sasaki, T.** (2005). Grains of knowledge: Genomics of model cereals. Genome Res **15**: 1643–1650.

**Pecinka, A., Abdelsamad, A., and Vu, G.T.H.** (2013). Hidden genetic nature of epigenetic natural variation in plants. Trends Plant Sci. **18**: 625–632.

**Pecinka, A., Chevalier, C., Colas, I., Kalantidis, K., Varotto, S., Krugman, T., Michailidis, C., Vallés, M.-P., Muñoz, A., and Pradillo, M.** (2020). Chromatin dynamics during interphase and cell division: similarities and differences between model and crop plants. J Exp Bot **71**: 5205–5222.

**Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., and Mittelsten Scheid, O.** (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell **22**: 3118–3129.

**Pecinka, A. and Mittelsten Scheid, O.** (2012). Stress-induced chromatin changes: a critical view on their heritability. Plant Cell Physiol. **53**: 801–808.

**Pecinka, A., Rosa, M., Schikora, A., Berlinger, M., Hirt, H., Luschnig, C., and Scheid, O.M.** (2009). Transgenerational stress memory is not a general response in Arabidopsis. PLoS One **4**.

**Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J., and Schubert, I.** (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. Chromosoma **113**: 258–269.

**Pellicer, J., Kelly, L.J., Leitch, I.J., Zomlefer, W.B., and Fay, M.F.** (2014). A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. New Phytol **201**: 1484–1497.

**Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L.** (2007). DNA demethylation in the Arabidopsis genome. Proc. Natl. Acad. Sci. **104**: 6752–6757.

**Picard, C.L., Povilus, R.A., Williams, B.P., and Gehring, M.** (2021). Transcriptional and imprinting complexity in Arabidopsis seeds at single-nucleus resolution. Nat. Plants **7**: 730–738.

**Pietzenuk, B., Markus, C., Gaubert, H., Bagwan, N., Merotto, A., Bucher, E., and Pecinka, A.** (2016). Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. Genome Biol. **17**: 1–15.

**Pilu, R.** (2015). Paramutation phenomena in plants. Semin. Cell Dev. Biol. **44**: 2–10.

**Przewloka, M.R. and Glover, D.M.** (2009). The Kinetochore and the Centromere: A Working Long Distance Relationship. Annu. Rev. Genet. **43**: 439–465.

**Qian, Y., Chen, C., Jiang, L., Zhang, J., and Ren, Q.** (2019). Genome-wide identification, classification and expression analysis of the JmjC domain-containing histone demethylase gene family in maize. BMC Genomics **20**: 256.

**Rangwala, S.H., Elumalai, R., Vanier, C., Ozkan, H., Galbraith, D.W., and Richards, E.J.** (2006). Meiotically stable natural epialleles of Sadhu, a novel Arabidopsis retroposon. PLoS Genet **2**: e36.

**Rangwala, S.H. and Richards, E.J.** (2010). The structure, organization and radiation of Sadhu non-long terminal repeat retroelements in Arabidopsis species. Mob. DNA **1**: 1–13.

**Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D.** (2006). N6-methyladenine: the other methylated base of DNA. Bioessays **28**: 309–315.

**Ravi, M., Shibata, F., Ramahi, J.S., Nagaki, K., Chen, C., and Chan, S.W.L.** (2011). Meiosis-Specific Loading of the Centromere-Specific Histone CENH3 in Arabidopsis thaliana. **7**.

**Rawat, V., Abdelsamad, A., Pietzenuk, B., Seymour, D.K., Koenig, D., Weigel, D., Pecinka, A., and Schneeberger, K.** (2015). Improving the annotation of Arabidopsis Lyrata using RNA-Seq data. PLoS One **10**: 1–12.

**Redman, J.C., Haas, B.J., Tanimoto, G., and Town, C.D.** (2004). Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. Plant J. **38**: 545–561.

**Rehrauer, H. et al.** (2010). AGRONOMICS1: A new resource for Arabidopsis transcriptome profiling. Plant Physiol **152**: 487–499.

**Reuter, J.A., Spacek, D. V, and Snyder, M.P.** (2015). High-throughput sequencing technologies. Mol Cell **58**: 586–597.

**Richards, E.J. and Ausubel, F.M.** (1988). Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. Cell **53**: 127–136.

**Riha, K., Heacock, M.L., and Shippen, D.E.** (2006). The Role of the Nonhomologous End-Joining DNA Double-Strand Break Repair Pathway in Telomere Biology. Annu. Rev. Genet. **40**: 237–277.

**Roa, F. and Guerra, M.** (2012). Distribution of 45S rDNA sites in chromosomes of plants: Structural and evolutionary implications. BMC Evol. Biol. **12**: 1–13.

**Rosa, S. and Shaw, P.** (2013). Insights into chromatin structure and dynamics in plants. Biology (Basel). **2**: 1378–1410.

**Roudier, F. et al.** (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J. **30**: 1928–1938.

**Rutowicz, K. et al.** (2015). A specialized histone H1 variant is required for adaptive responses to complex abiotic stress and related DNA methylation in Arabidopsis. Plant Physiol. **169**: 2080–2101.

**Sakai, H., Mizuno, H., Kawahara, Y., Wakimoto, H., Ikawa, H., Kawahigashi, H., Kanamori, H., Matsumoto, T., Itoh, T., and Gaut, B.S.** (2011). Retrogenes in rice (Oryza sativa L. ssp. japonica) exhibit correlated expression with their source genes. Genome Biol. Evol. **3**: 1357–1368.

**Sakurai, H. and Enoki, Y.** (2010). Novel aspects of heat shock factors: DNA recognition, chromatin modulation and gene expression. FEBS J. **277**: 4140–4149.

**Schranz, E.M., Mohammadin, S., and Edger, P.P.** (2012). Ancient whole genome duplications, novelty and diversification: The WGD Radiation Lag-Time Model. Curr. Opin. Plant Biol. **15**: 147–153.

**Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierreza, C.** (2014). The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. Plant Cell **26**: 2351–2366.

**Seymour, D.K., Koenig, D., Hagmann, J., Becker, C., and Weigel, D.** (2014). Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. PLoS Genet. **10**.

**Shaked, H., Avivi-Ragolsky, N., and Levy, A.A.** (2006). Involvement of the arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and

recombination. Genetics **173**: 985–994.

**De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C.E., Maere, S., and Van de Peer, Y.** (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U.S.A. **110**: 2898–2903.

**Springer, N.M., Li, Q., and Lisch, D.** (2016). Creating order from chaos: epigenome dynamics in plants with complex genomes. Plant Cell **28**: 314–325.

**Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y. V, and Jacobsen, S.E.** (2012a). Resource comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell **152**: 352–364.

**Stroud, H., Hale, C.J., Feng, S., Caro, E., Jacob, Y., Michaels, S.D., and Jacobsen, S.E.** (2012b). DNA methyltransferases are required to induce heterochromatic re-replication in arabidopsis. PLoS Genet. **8**: 1–7.

**Takayama, S., Dhahbi, J., Roberts, A., Mao, G., Heo, S.-J., Pachter, L., Martin, D., and Boffelli, D.** (2014). Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. Genome Res.

**Takuno, S. and Gaut, B.S.** (2012). Recent Retrotransposon Insertions Are Methylated and Phylogenetically Clustered in Japonica Rice ( Oryza sativa spp . japonica ). **29**: 3193–3203.

**Tenaillon, M.I., Hollister, J.D., and Gaut, B.S.** (2010). A triptych of the evolution of plant transposable elements. Trends Plant Sci **15**: 471–478.

**Thanbichler, M. and Shapiro, L.** (2006). Chromosome organization and segregation in bacteria. J. Struct. Biol. **156**: 292–303.

**Thomas, C.A.J.** (1971). The genetic organization of chromosomes. Annu Rev Genet **5**: 237–256.

**Tiang, C., He, Y., and Pawlowski, W.P.** (2012). Chromosome Organization and Dynamics during Interphase , Mitosis , and Meiosis in Plants 1. **158**: 26–34.

**Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant, I.** (2010). Stress-induced activation of heterochromatic transcription. PLoS Genet. **6**: 1–11.

**Tsukahara, S., Kobayashi, A., Kawabe, A., Mathieu, O., Miura, A., and Kakutani, T.** (2009). Bursts of retrotransposition reproduced in Arabidopsis. Nature **461**: 423–426.

**Vanyushin, B.F., Alexandrushkina, N.I., and Kirnos, M.D.** (1988). N6-Methyladenine in mitochondrial DNA of higher plants. FEBS Lett. **233**: 397–399.

**Vu, G.T.H. et al.** (2015). Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. Plant Genome **8**: doi:10.3835/plantgenome2015.04.0021.

**Waddington, C.H.** (1942). The epigenotype. Endeavour: 18–20.

**Wang, W. et al.** (2006a). High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell **18**: 1791–1802.

**Wang, W. et al.** (2006b). High Rate of Chimeric Gene Origination by Retroposition in Plant Genomes. **18**: 1791–1802.

**Waterland, R.A. and Jirtle, R.L.** (2003). Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. Mol. Cell. Biol. **23**: 5293–5300.

**Watson, J.D. and Crick, F.H.C.** (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature **171**: 737–738.

**Watson, J.M. and Riha, K.** (2010). Comparative biology of telomeres: Where plants stand. FEBS Lett. **584**: 3752–3759.

**Wicker, T. et al.** (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet **8**: 973–982.

**Wicker, T. et al.** (2017). The repetitive landscape of the 5100 Mbp barley genome. Mob. DNA **8**: 22.

**Willing, E.M. et al.** (2015). Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. Nat. Plants **1**: 1–7.

**Willing, E.M., Piofczyk, T., Albert, A., Winkler, J.B., Schneeberger, K., and Pecinka, A.** (2016). UVR2 ensures transgenerational genome stability under simulated natural UV-B in Arabidopsis thaliana. Nat. Commun. **7**: 1–9.

**Woo, H.R., Dittmer, T.A., and Richards, E.J.** (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in arabidopsis. PLoS Genet. **4**.

**Woo, H.R., Pontes, O., Pikaard, C.S., and Richards, E.J.** (2007). protein required for centromeric heterochromatinization. **1**: 267–277.

**Xiong, W., He, L., Lai, J., Dooner, H.K., and Du, C.** (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc Natl Acad Sci U.S.A. **111**: 10263–10268.

**Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V.** (2009). Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. Proc. Natl. Acad. Sci. **106**: 4941 LP – 4946.

**Yang, F., Fernández-Jiménez, N., Tučková, M., Vrána, J., Cápal, P., Díaz, M., Pradillo, M., and Pecinka, A.** (2021). Defects in meiotic chromosome segregation lead to unreduced male gametes in Arabidopsis SMC5/6 complex mutants. Plant Cell.

**Yang, L. and Bennetzen, J.L.** (2009). Structure-based discovery and description of plant and animal Helitrons. Proc. Natl. Acad. Sci. U. S. A. **106**: 12832–12837.

**Yelagandula, R. et al.** (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in arabidopsis. Cell **158**: 98–109.

**Yoshida, S., Maruyama, S., Nozaki, H., and Shirasu, K.** (2010). Horizontal Gene Transfer by the Parasitic Plant <em>Striga hermonthica</em>. Science (80-. ). **328**: 1128.

**Zakrzewski, F., Schubert, V., Viehoever, P., Minoche, A.E., Dohm, J.C., Himmelbauer, H., Weisshaar, B., and Schmidt, T.** (2014). The CHH motif in sugar beet satellite DNA: A modulator for cytosine methylation. Plant J. **78**: 937–950.

**Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D.** (2013). The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell **153**: 193–205.

**Zhang, Y., Wu, Y., Liu, Y., and Han, B.** (2005). Computational identification of 69 retroposons in Arabidopsis. Plant Physiol **138**: 935–948.

**Zhu, J.-K.** (2009). Active DNA Demethylation Mediated by DNA Glycosylases. Annu. Rev. Genet. **43**: 143–166.

**Zilberman, D., Coleman-derr, D., Ballinger, T., and Henikoff, S.** (2008). Histone H2A . Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456**: 125–130.

# 7. PODĚKOVÁNÍ

## 8. SEZNAM PRACÍ TVOŘÍCÍCH HABILITAČNÍ SPIS

1. **Pecinka A.**, Chevalier C., Colas I., Kalantidis K., Varotto S., Krugman T., Michailidis C., Vallés M.-P., Muñoz A., Pradillo M. (2020): Chromatin dynamics during interphase and cell division: similarities and differences between model and crop plants. *J Exp Bot* 71: 5205–5222.

2. Nowicka A., Tokarz B., Zwyrtková J., Dvořák Tomaštíková E., Procházková K., Ercan U., Finke A., Rozhon W., Poppenberger B., Otmar M., Niezgodzki I., Krečmerová M., Schubert I., **Pecinka A.*** (2020): Comparative analysis of epigenetic inhibitors reveals different degrees of interference with transcriptional gene silencing and induction of DNA damage. *Plant J* 102:68-84.

3. Díaz M., Pecinkova P., Nowicka A., Baroux C., Sakamoto T., Gandha P.Y., Jeřábková H., Matsunaga S., Grossniklaus U., *****Pecinka A.** (2019): SMC5/6 complex subunit NSE4A is involved in DNA damage repair and seed development in Arabidopsis. *Plant Cell* 31:1579-1597.

4. Finke A., Mandáková T., Nawaz K., Vu G.T.H., Novák P., Macas J., Lysak M.A., **Pecinka A.** (2019): Genome invasion by a hypomethylated satellite repeat in Australian crucifer *Ballantinia antipoda*. *Plant J* 99:1066-1079.

5. Willing E-M, Piofczyk T, Albert A, Winkler JB, Schneeberger K, **Pecinka A** (2016): UVR2 ensures trans-generational genome stability under simulated natural UV-B in *Arabidopsis thaliana*. *Nature Communications*, 7: e13522.

6. Pietzenuk B., Markus C., Gaubert H., Bagwan N., Abdelsamad A., Merotto A., Bucher E., **Pecinka A.** (2016): Recurrent evolution of heat-responsiveness in *Brassicaceae COPIA* elements. *Genome Biology*, 17: e209.

7. Liu C.-H., Finke A., Díaz A., Rozhon W., Poppenberger B., Baubec T., **Pecinka A.** (2015): ATR and ATM are required for repair of zebularine-induced DNA damage in *Arabidopsis thaliana*. *Plant Cell* 27: 1788-1800.

8. Abdelsamad A, **Pecinka A** (2014): Pollen-specific activation of Arabidopsis retrogenes is associated with global transcriptional reprogramming. *Plant Cell* 26: 3299-313.

9. Baubec T, Finke A, Mittelsten Scheid O, *****Pecinka A** (2014): Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Rep* 15: 446-452.

10. **Pecinka A**, Mittelsten Scheid O (2012): Stress-induced chromatin changes: A critical view on their heritability. *Plant Cell Physiol* 53: 801–808.

11. Baubec T, Dinh HQ, **Pecinka A**, Rakic B, Rozhon W, Wohlrab B, von Haeseler A, Mittelsten Scheid O (2010): Cooperation of Multiple Chromatin Modifications Can Generate Unanticipated Stability of Epigenetic States in Arabidopsis. *Plant Cell* 22: 34-47.

12. **Pecinka A**, Dinh HQ, Rosa M, Baubec T, Lettner N, Mittelsten Scheid O (2010): Epigenetic control of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell* 22: 3118–3129.

13. Baubec T, **Pecinka A**, Rozhon W, Mittelsten Scheid O (2009): Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *Plant J* 57:542-554.

# PŘÍLOHA: PUBLIKACE TVOŘÍCÍ HABILITAČNÍ SPIS

# Publikace 1
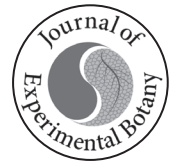
REVIEW PAPER

# Chromatin dynamics during interphase and cell division: similarities and differences between model and crop plants

**Ales Pecinka[1],[*],** [ID] **, Christian Chevalier[2],** [ID] **, Isabelle Colas[3],** [ID] **, Kriton Kalantidis[4],** [ID] **, Serena Varotto[5],** [ID] **, Tamar Krugman[6],** [ID] **, Christos Michailidis[7],** [ID] **, María-Pilar Vallés[8], Aitor Muñoz[9] and Mónica Pradillo[10],** [ID]

[1] Institute of Experimental Botany, Czech Acad Sci, Centre of the Region Haná for Agricultural and Biotechnological Research, Šlechtitelů 31, Olomouc, CZ-779 00, Czech Republic

[2] UMR1332 BFP, INRA, University of Bordeaux, 33882 Villenave d'Ornon Cedex, France

[3] James Hutton Institute, Cell and Molecular Science, Pr Waugh's Lab, Errol Road, Invergowrie, Dundee DD2 5DA, UK

[4] Department of Biology, University of Crete, and Institute of Molecular Biology Biotechnology, FoRTH, Heraklion, 70013, Greece

[5] Department of Agronomy Animal Food Natural Resources and Environment (DAFNAE) University of Padova, Agripolis viale dell'Università, 16 35020 Legnaro (PD), Italy

[6] Institute of Evolution, University of Haifa, Abba Khoushy Ave 199, Haifa, 3498838, Israel

[7] Institute of Experimental Botany, Czech Acad Sci, Rozvojová 263, Praha 6 - Lysolaje, CZ-165 02, Czech Republic

[8] Department of Genetics and Plant Breeding, Estación Experimental Aula Dei (EEAD), Spanish National Research Council (CSIC), Zaragoza, Spain

[9] Department of Plant Molecular Genetics, National Center of Biotechnology/Superior Council of Scientific Research, Autónoma University of Madrid, 28049 Madrid, Spain

[10] Department of Genetics, Physiology and Microbiology, Faculty of Biology, Complutense University of Madrid, 28040 Madrid, Spain

* Correspondence: pecinka@ueb.cas.cz

## Abstract

**Genetic information in the cell nucleus controls organismal development and responses to the environment, and finally ensures its own transmission to the next generations. To achieve so many different tasks, the genetic information is associated with structural and regulatory proteins, which orchestrate nuclear functions in time and space. Furthermore, plant life strategies require chromatin plasticity to allow a rapid adaptation to abiotic and biotic stresses. Here, we summarize current knowledge on the organization of plant chromatin and dynamics of chromosomes during interphase and mitotic and meiotic cell divisions for model and crop plants differing as to genome size, ploidy, and amount of genomic resources available. The existing data indicate that chromatin changes accompany most (if not all) cellular processes and that there are both shared and unique themes in the chromatin structure and global chromosome dynamics among species. Ongoing efforts to understand the molecular mechanisms involved in chromatin organization and remodeling have, together with the latest genome editing tools, potential to unlock crop genomes for innovative breeding strategies and improvements of various traits.**

**Keywords:** Arabidopsis, chromatin, chromosome, crops, epigenetics, mitosis, meiosis, plant breeding, plant development.

## Introduction

Most eukaryotic DNA, the carrier of genetic information, is stored in cell nuclei as linear supermolecules—the chromosomes. Complexes of nuclear DNA with the associated proteins constitute chromatin, which is required for proper DNA packaging, regulation of gene expression, and chromosome organization. The basic units of chromatin are the nucleosomes, which consist of ~146 bp of DNA wrapped around a histone octamer having two copies of each of H2A, H2B, H3, and H4 (reviewed in, for example, McGinty and Tan, 2015).

Replacing the canonical histones with non-canonical ones leads to different chromatin functions (Koyama and Kurumizaka, 2018). Data from the model species *Arabidopsis thaliana* (Arabidopsis) suggest functional diversification of histone H1, H2A, and H3 proteins. Histones H1.1 and H1.2 represent the canonical forms, but H1.3 is a stress–inducible variant (Rutowicz *et al.*, 2015). The H2A.Z-containing nucleosomes occur in the transcription start and termination sites of ubiquitously transcribed genes and cover large parts of stress– and developmentally regulated genes (Coleman-Derr and Zilberman, 2012). H2A.Z also marks other functional domains, such as potential sites of meiotic recombination (Zilberman *et al.*, 2008; Choi *et al.*, 2013; Yelagandula *et al.*, 2014). H2A.X is an evolutionarily conserved variant scattered throughout the genome and, upon phosphorylation of the Ser139 residue (γ-H2A.X), labels the sites of DNA damage repair (Friesner *et al.*, 2005; Lorković *et al.*, 2017). The recently discovered plant-specific variant H2A.W occurs in repetitive DNA regions, where it represses transposons and marks the sites of DNA damage repair (Yelagandula *et al.*, 2014; Lorković *et al.*, 2017). The H3 proteins include H3.1, H3.3, and CenH3 (CENP-A), representing the transcriptionally active, inactive, and the kinetochore-binding regions, respectively (Lermontova *et al.*, 2011; Stroud *et al.*, 2012; Wollmann *et al.*, 2012; Maheshwari *et al.*, 2015). CenH3 receives a good deal of attention owing to the fact that its mutations lead to production of haploids, a trait that could be used in the process of double haploid production (Ravi and Chan, 2010; Sanei *et al.*, 2011; Yuan *et al.*, 2015; Karimi-Ashtiyani *et al.*, 2015).

Unstructured histone N-termini (tails) are the rich substrate for post-translational modifications (PTMs) by methylation, acetylation, and phosphorylation, among others. Acetylation is associated with active chromatin, while methylation can have both permissive and repressive functions depending on the residue and the number of methyl groups in plants.

The most common plant genome DNA modification is cytosine methylation (5-methyl-2'-deoxy-cytosine or DNA methylation), where CG, CHG, and CHH (H=A, T, or C) represent the three functional DNA methylation contexts (Law and Jacobsen, 2010). DNA methylation can be established *de novo* at any cytosine by the RNA-directed DNA methylation (RdDM) pathway guided to the target sequences by siRNAs with perfect sequence homology (reviewed in, for example, Matzke and Mosher, 2014). Once established, DNA methylation is maintained by the activity of replication–coupled DNA methyltransferases specialized for each cytosine context, and by the corrective activity of RdDM (Du *et al.*, 2012; Zemach

*et al.*, 2013; Baubec *et al.*, 2014). So far, little is known about the significance and the functions of adenine methylation in plants (Vanyushin *et al.*, 1988; Fu *et al.*, 2015).

Nucleosomal DNA arrays are folded at multiple levels into higher order structures and eventually into the chromosomes (reviewed in, for example, Dixon *et al.*, 2016). Microscopic observations of variable chromatin staining intensity led to the early description of the darker chromosome stain called heterochromatin and the lighter chromosome stain called euchromatin (Heitz, 1928). Molecular experiments revealed that heterochromatin is normally repeat rich/gene poor, densely packed, and transcriptionally silent, while euchromatin is open, repeat poor/gene rich, and transcriptionally active (Roudier *et al.*, 2011; Sequeira-Mendes *et al.*, 2014). The organization and dynamics of the large chromatin domains and their functional significance in plants seem to be strongly influenced by the nuclear genome size and amount of repetitive DNA, but it is still not well understood. The small genome of Arabidopsis is organized as mostly randomly positioned chromosome territories with nuclear envelope (NE)-associated heterochromatic chromocenters (CCs) and nucleolus–associated nucleolar organizer regions (NORs) (Fransz *et al.*, 2002; Pecinka *et al.*, 2004). In contrast, large genomes of cereals, for example, show Rabl organization with centromeres and telomeres clustered at the opposite poles of the nuclei. These patterns have recently been explored in detail by the chromatin conformation capture techniques (reviewed in Doğan and Liu, 2018). Currently, it remains unknown how representative such organizations are for different tissues, under changing environmental conditions, and for species with intermediate DNA content. In addition, Hi-C (high-throughput chromosome conformation capture) experiments suggest that a combination of different factors, such as genomic composition, epigenetic modification, and transcriptional activity, are involved in shaping global and local chromatin packing in Arabidopsis and rice (Grob *et al.*, 2014; Dong *et al.*, 2018). Hi-C applications to other crops will improve our knowledge of the role of chromosomes packing in the nucleus in modifying gene expression under stress conditions.

## Chromatin organization in somatic cell nuclei under ambient and stress conditions

Plants rapidly change gene expression during stress, to make a rational use of the existing resources and to minimize damage. Chromatin changes have been found after practically all types of applied abiotic and biotic stresses, and there is growing evidence that some epigenetic changes play an important role in the fine-tuning of stress responses (Kim *et al.*, 2010; Ding and Wang, 2015) (Fig. 1).

Nuclei of germinating Arabidopsis seeds appear mostly euchromatic, and heterochromatin is established only in response to the light stimulus (Mathieu *et al.*, 2003). Light-induced heterochromatin re-organization leads to transcriptional reprogramming and activation of photosynthesis during germination
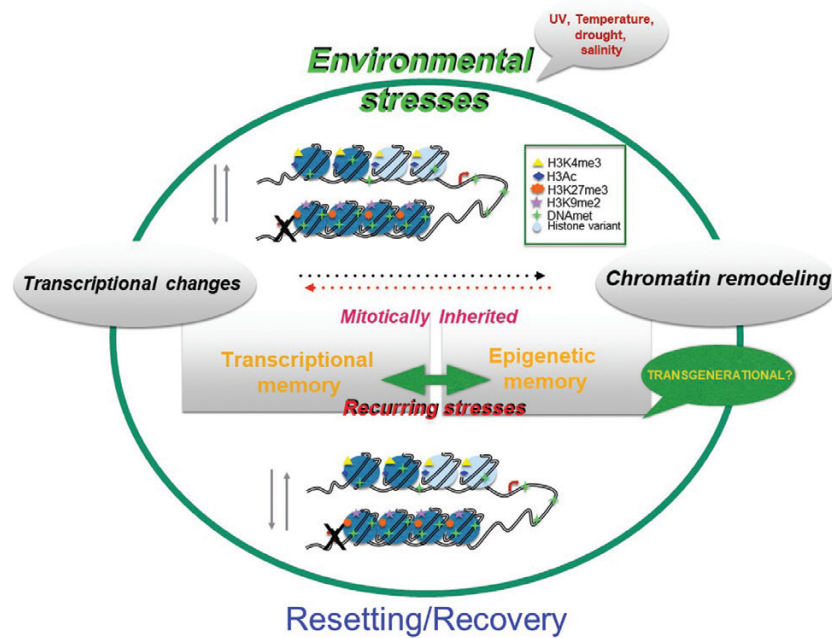
**Fig. 1.** Overview of stress-induced chromatin changes and their potential trajectories. Environmentally induced stresses lead to genome-wide changes of transcript levels. These changes are accompanied by dynamic changes influencing chromatin compaction and also gene expression. Transcriptional and chromatin changes can be correlated or uncorrelated, and the exact hierarchy of events determining these changes can vary according to the plant species and type of stress. There is some evidence that both transcriptional and chromatin changes can persist after the removal of stress and can be mitotically inherited. In a transcriptional memory gene, high expression levels persistent for a prolonged period of time even after the end of a stress cue. In the case of recurring stress, the transcriptional response to a second stress cue is modified compared with the response to the first exposure to the same stress. Many cases of memory also involve chromatin dynamics at key regulatory loci (epigenetic memory). Despite transcriptional and chromatin/epigenetic memory, resetting and recovery are probably the over-riding strategies used by plants to maximize fitness in time and space.

(Bourbousse *et al.*, 2015). Light quality-induced phytochrome signaling may also cause repositioning of specific chromatin regions, such as the chlorophyll A/B binding (CAB) locus in Arabidopsis, and thus influence gene expression (Feng *et al.*, 2014). The composition and intensity of solar radiation varies strongly depending on the season, geographical location, or surrounding vegetation.

UV A and B (UV-A/B, 280–400 nm) is the most energetic component of solar radiation, which damages membranes, proteins, and DNA, and its intensity increases with altitude and proximity to the equator. Plants probably adapt to UV radiation as indicated by the constitutive expression of chromatin-remodeling factors and reduced sensitivity to UV damage, as was found in maize landraces at tropical high altitude (Casati *et al.*, 2006, 2008). Interestingly, methyl cytosines have a higher propensity to be involved in UV-induced pyrimidine dimers than normal cytosines, and their less efficient repair in heterochromatin leads to conversions into thymines (Willing *et al.*, 2016). Hence, UV radiation has a profound effect on both epigenome and genome stability.

Temperature fluctuations are common and involve rapid adjustment of cellular metabolism, growth, and differentiation (Kotak *et al.*, 2007). Heat stress reduces chromatin compaction and the coordinated organ-specific transcriptional response via changes in nucleosome and H2A.Z occupancy (Kumar and Wigge, 2010; Pecinka *et al.*, 2010; Boden *et al.*, 2013; Lämke and Bäurle, 2017). Severe heat stress modulates chromatin structure, by increasing histone acetylation and decreasing H3K9me2,

and eventually induces programmed cell death (Z. Wang *et al.*, 2015). Surprisingly, cold stress also leads to general chromatin de-condensation, as suggested by Hi–C analysis in rice, but specific regions may be subject to chromatin condensation and gene silencing (Liu *et al.*, 2017). Taken together, the data suggest that at a range of optimal temperatures, which are species specific, chromatin is normally condensed, and de-condenses under suboptimal conditions. However, this hypothesis needs to be tested for a broader range of species and temperatures.

Vernalization—acquisition of competence to flower only in response to a period of cold—is a well-known example of cold-induced chromatin change. In Arabidopsis, vernalization occurs via H3K27 tri-methylation and silencing of the MADS box transcription repressor *FLOWERING LOCUS C* (*FLC*) (Rosa and Shaw, 2013; Whittaker and Dean, 2017). However, vernalization evolved multiple times in plants and its mechanism differs between species (Reeves *et al.*, 2012; Périlleux *et al.*, 2013; Ruelens *et al.*, 2013; Porto *et al.*, 2015). *VERNALIZATION 1* (*VRN1*) is the major vernalization gene in cereals, which loses H3K27me3 and gains H3K4me3 during cold periods (Oliver *et al.*, 2009; Diallo *et al.*, 2012). Temperature changes also lead to selective and transient activation of repetitive sequences (Steward *et al.*, 2002; Pecinka *et al.*, 2010; Tittel-Elmer *et al.*, 2010; Ito *et al.*, 2011). Recent studies suggested that this is due to the presence of the canonical *cis*-regulatory elements in the long terminal repeats (LTRs) of specific stress-responsive transposon families (Cavrak *et al.*, 2014; Pietzenuk *et al.*, 2016). This could represent an evolutionary

mechanism of dispersal for *cis*-regulatory elements in the genome and foundation of novel gene expression patterns (Ito *et al.*, 2011).

Reduced water availability negatively influences yield and resistance to other stresses. The effect of water stress on plant chromatin is not well understood, but data suggest that the responses are species specific. Drought caused DNA methylation changes in the shoot apical meristems (SAMs) of hybrid poplars (Gourcilleau *et al.*, 2010), and there were additional changes in DNA methylation and expression of phytohormone metabolism genes after re-watering (Gourcilleau *et al.*, 2010). In tomato, drought-induced DNA methylation changes in *ABSCISIC ACID STRESS AND RIPENING 1* and *2* (*ASR1* and *ASR2*) genes (González *et al.*, 2011, 2013), and thus probably modified the ripening process.

In contrast, no consistent water stress-induced DNA methylation changes were observed in Arabidopsis and maize (Eichten and Springer, 2015; Ganguly *et al.*, 2017). Instead, H3K4me3 may represent a drought stress 'memory' mark, which influences the transcriptional response during recurring stress in Arabidopsis (Ding *et al.*, 2012). The topic of chromatin-mediated 'epigenetic memory' has been recently reviewed in several papers (for example, in Jablonka and Raz, 2009; Avramova, 2015; Lämke and Bäurle, 2017), and therefore we do not review it here.

Attacks of crops by pathogens may have severe consequences on plant vitality and yield, and can even cause lethality. Biotic stress defense mechanisms are fast evolving to match the evolutionary innovations on the pathogen side, which leads to a constant race between the host and the pathogen. Following infection by biotrophic or necrotrophic pathogens, plants typically reprogram gene expression from growth to defense (Moore *et al.*, 2011), which involves activation of the salicylic acid (SA) and the jasmonic acid/ethylene (JA/ET) pathways, respectively (reviewed in, for example, Glazebrook, 2005; Vlot *et al.*, 2009). Some pathogens developed strategies to directly affect chromatin modifiers. For example, the necrotrophic fungus *Alternaria brassicola* produces a toxin that inhibits the enzyme histone deacetylase (HDA) activity during infection (Matsumoto *et al.*, 1992; Kwon *et al.*, 2003). In line with this, knockdown of Arabidopsis HDA19 led to increased susceptibility to *A. brassicola*, while its overexpression activated JA/ET-regulated genes and triggered pathogen resistance (Zhou *et al.*, 2005). HDA19 represses SA biosynthesis and defense responses in Arabidopsis by suppressing transcription of *PATHOGENESIS RELATED* (*PR*) *PR1* and *PR5* genes (Tian *et al.*, 2005), indicating its negative role in SA-mediated defense responses (Choi *et al.*, 2012). Upon infection by *Pst* DC3000, *SIRTUIN2* (*SRT2*), another HDA involved in immune responses, is down-regulated, leading to higher SA production and expression of downstream defense genes (Wang *et al.*, 2010). In contrast, some HDAs regulate innate immunity positively (Latrasse *et al.*, 2017a). Although it is clear that histone acetylation (and de-acetylation) plays an important role in the regulation of defense-related genes, it is still not clear how HAT and HDAs are targeted to the target loci to allow genome-wide changes in gene expression (Ramirez-Prado *et al.*, 2018).

The effects of viruses on plant chromatin remain only poorly understood. In a pioneer study, Arabidopsis mutants deficient in DNA methylation and RdDM were found to be susceptible to geminiviruses (Raja *et al.*, 2008). The genimivirus genome consists of two ssDNA molecules, which replicate using the host's replication machinery. The replicated virus dsDNAs are packed with nucleosomes and form tiny chromosome-like structures. The hosts' defense responses involve suppression of gene expression by methylating the viral genome. Involvement of RNA polymerase II (Pol II) and RDR6 (Jackel *et al.*, 2016) indicates that the silencing is triggered by the non-canonical RdDM (reviewed in, for example, Matzke and Mosher, 2014).

In summary, this section shows that responses of chromatin to various stresses are diverse and in some cases highly adaptive. In many cases, we have only a basic description of the stress-induced chromatin changes, and we are still lacking information on the persistence of these changes after recovery from the stress and about their heritability through mitosis and meiosis. Therefore, we expect that many future studies will focus on the identification of the underlying mechanisms. In addition, it is expected that more groups of chromatin modifiers such as histone (de)methyltransferases and (de)ubiquitinylases will be firmly connected with stress-induced chromatin responses (Dhawan *et al.*, 2009; L.C. Wang *et al.*, 2015; Dutta *et al.*, 2017). Understanding the involvement of chromatin in adjusting plant adaptation to diverse environmental challenges is of interest to a broad audience of plant scientists, considering that stresses are generally predicted to become exacerbated due to climate change and that they can strongly affect crop yields.

## Chromatin organization during mitotic and meiotic cell divisions

Chromatin undergoes drastic changes affecting its degree of compaction during the cell cycle. At the onset of cell divisions, the NE disassembles, allowing the access of cytoplasmic proteins to the nucleoplasm, including proteins which contribute to further chromatin condensation and spindle formation. Chromatin condensation is critical for the individualization of chromosome in order to guarantee the proper distribution of genetic information between daughter cells. After segregation, chromatin is decondensed to restore its interphase state. To achieve this process, specific PTMs in histones occur, including the marker of condensed chromatin, histone H3S10p (p=phosphorylation), and mitosis-specific PTMs such as histone H3T3p and H3T11p (Houben *et al.*, 2002; Zhang *et al.*, 2005). In maize, histone H3S28p and H3S50p delineate the pericentromeric and centromeric regions during chromosome segregation, respectively (Zhang *et al.*, 2005). In the same species, changes in the level of histone H3S10p regulate sister chromatid cohesion (Kaszas and Cande, 2000), and an increase of H3 phosphorylation is linked to reduced acetylation levels at Lys9 residues in histone H3 (Edmondson *et al.*, 2002). In barley, histone H4 acetylation (K5, K8, K12, and K16) is an important modification for chromatin structure, with H4K8Ac having no impact on chromatin structure from mitotic prophase to telophase (similar to H4K16Ac), while H4K5Ac

and H4K12Ac are more dynamic (Wako *et al.*, 2003, 2005). A survey of 17 plants species revealed that the distribution of histone H4K5ac differs between small and large genome species (Feitoza *et al.*, 2017). In most small genome species (2C <5 pg), H4K5ac was enriched in late condensing terminal regions but depleted in early condensing regions, while in large genome species, acetylation was more evenly displayed across the chromosomes which were also uniformly condensed during the prophase stage.

The condensin complex is another main player in chromosome organization (Hirano *et al.*, 1997), which is probably recruited by H3S10p (Schmiesing *et al.*, 2000). Its basic structure is given by the heterodimer of structural maintenance of chromosomes (SMC) proteins SMC2 and SMC4, with which condensin I- and II-specific regulatory subunits associate. Condensin II accesses the cell nucleus before mitosis and its reduction partially reduces early H3 phosphorylation (Ono *et al.*, 2004). Subsequently, condensin I contributes to prophase chromatin compaction.

Similarly, the cohesin complex also contains two SMC subunits (SMC1 and SMC3), that are connected by an α-kleisin subunit (represented by one of the four homologs SYN1–SYN4 in Arabidopsis), which recruits the HEAT repeat–containing subunit SCC3. In addition, different proteins regulate cohesion establishment and maintenance (Bolaños-Villegas *et al.*, 2017). Cohesion is established at the onset of S phase and persists until the metaphase–anaphase transition, and it is essential to resist the force of the spindle microtubules while chromosomes are aligned at the equatorial plate, allowing their accurate segregation to opposite poles (Fig. 2). At the beginning of anaphase, cohesin is released from chromosomes in two steps (Nasmyth, 2001). During prophase and prometaphase, cohesin is removed from chromosome arms. In the second step, before the onset of anaphase, the remaining cohesin is released from centromeres, allowing separation of sister chromatids. The PRECOCIOUS DISSOCIATION OF SISTERS 5–WING APART LIKE (PDS5–WAPL) complex eliminates cohesin from chromosome arms, whereas EXTRA SPINDLE POLE BODIES 1 (ESP1) separase removes centromeric cohesin via an ubiquitin-dependent cleavage of the α-kleisin in Arabidopsis (Liu and Makaroff, 2006; Pradillo *et al.*, 2015; De *et al.*, 2016). ESP1 is also important for the proper establishment of the radial microtubule network and nuclear/cytoplasmic domains (Yang *et al.*, 2009). Several studies have demonstrated that cohesin plays additional roles in DNA double-strand break repair (DSBR) and regulation of gene expression (Yuan *et al.*, 2011; Mehta *et al.*, 2012).

There are remarkable differences in chromatin condensation and organization between mitosis and meiosis (Fig. 2). Meiotic chromosome condensation proceeds simultaneously with alignment of homologous chromosomes, programmed DSB formation, repair through homologous recombination (HR), and establishment and dissolution of the synaptonemal complex (SC). These processes are associated with striking morphological changes including dynamic variations in histone PTMs (Nasuda *et al.*, 2005; Oliver *et al.*, 2013). In leptonema, sites of DSB formation and their repair become marked with γ-H2A.X (Shroff *et al.*, 2004). In pachynema, γ-H2A.X is completely lost from fully synapsed chromosomes. In barley, the first γ-H2 A.X foci appeared only 4 h after DNA replication in pollen mother cells (PMCs) (Higgins *et al.*, 2012; He *et al.*, 2017). In Arabidopsis, DSB hotspots are also associated with the markers of active chromatin, including the histone H2A.Z variant and H3K4me3 modification, low nucleosome density, and low DNA methylation (Choi *et al.*, 2013). Similarly, crossovers (COs) reside in genomic regions of 'open chromatin', which were identified based on hypersensitivity to DNase I digestion and H3K4me3–enriched nucleosomes in potato (Marand *et al.*, 2017). This is also likely to be the case for barley as DSBs and H3K4me3 are strongly localized towards the telomeres, whereas they are quite low in pericentromeric regions (Baker *et al.*, 2015). However, only 20% of the DSBs are effectively associated with H3K4me3, leaving the other 80% unexplained in maize (Sidhu *et al.*, 2015; He *et al.*, 2017).

SWITCH1 (SWI1) is a plant-specific protein that regulates the switch from mitosis to meiosis (Mercier *et al.*, 2001; Agashe *et al.*, 2002; Sheehan and Pawlowski, 2009). Recently, it has been reported that SWI1 antagonizes WAPL during prophase I through a Sororin-like strategy in mitosis (Yang *et al.*, 2019). *swi1* mutants have altered distribution of acetylated histone H3 and dimethylated histone H3 (H3K4me2) (Boateng *et al.*, 2008). Interestingly, H3K4me2 is recognized by MALE MEIOCYTE DEATH 1 (MDD1), a PHD finger protein which acts as a transcriptional regulator, essential for Arabidopsis male meiosis (Andreuzza *et al.*, 2015). Arabidopsis plants defective for ARABIDOPSIS SKP1-LIKE1 (ASK1), a component of the SKP1–CUL1–F-box (SCF) ubiquitin ligase, also displays variations in acetylated histone H3 and H3K9me2 distribution patterns during meiosis (Yang *et al.*, 2006). The influence of these PTMs in meiotic HR has been highlighted in a recent work in which the disruption of H3K9me2 and DNA methylation pathways produces the epigenetic activation of meiotic recombination near centromeres (Choi *et al.*, 2018; Underwood *et al.*, 2018). These are regions normally suppressed for COs in order to avoid aneuploidies in the offspring (Rockmill *et al.*, 2006). In rice, the chromosomes are reprogrammed during the transition to meiosis under the control of the Argonaute protein MEIOSIS ARRESTED AT LEPTOTENE 1 (MEL1), increasing H3K9me2 and decreasing H3K9ac and H3S10p in order to promote synapsis and HR (Liu and Nonomura, 2016).

Entangling of meiotic prophase I chromosomes results in interlocks (Gelei, 1921), which could compromise chromatin integrity and result in chromosome mis-segregation. Here, the organization and movements of chromosome termini (typically traced by labeling of telomeric repeats) and TOPOISOMERASE II (TOPII) activity are essential for removal of the interlocks (Martinez-Garcia *et al.*, 2018). At the onset of meiosis, telomeres attach to the NE and cluster, forming a characteristic bouquet arrangement (Bass *et al.*, 2000). The mechanism of bouquet formation is not well understood and, although it is widely conserved among eukaryotes, a characteristic bouquet arrangement is apparently not formed in Arabidopsis (Armstrong *et al.*, 2001). In Arabidopsis, telomeres present a complex behavior and are associated with the nucleolus throughout meiotic interphase and
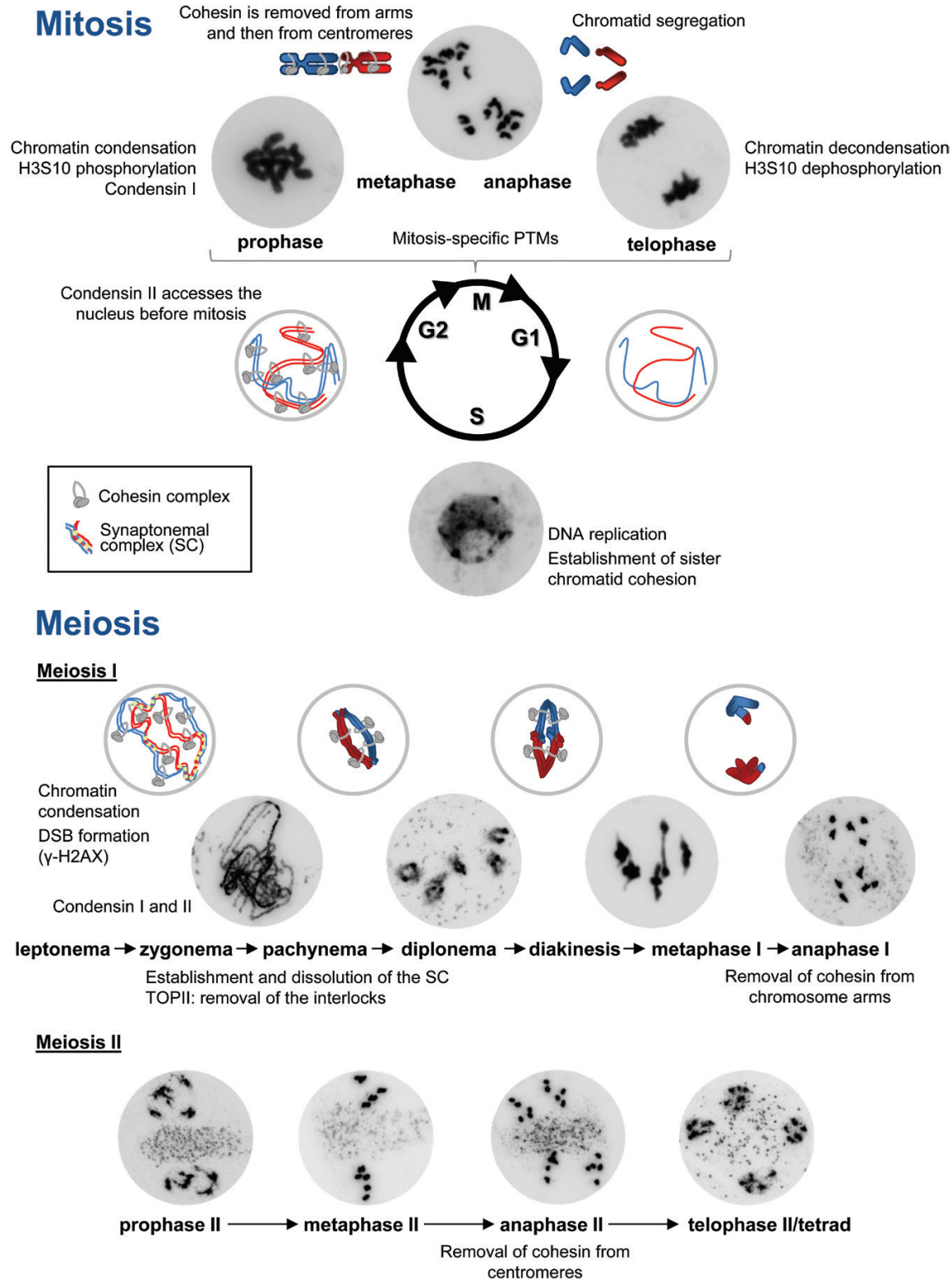
**Fig. 2.** Overview of chromosome organization during mitosis and meiosis. At the onset of mitosis, chromatin condensation is necessary to disassemble the interphase chromatin in a process driven by specific post-translational modifications (PTMs) in H3 and condensin complexes. In addition, the cohesin complex is essential for defining chromosome structure by providing a physical linkage between sister chromatids until their segregation at anaphase. Throughout meiosis, condensin complexes I and II are required to maintain the structural integrity of chromosomes. During leptonema, the histone variant H2A.X is rapidly phosphorylated to γ-H2A.X at double-strand break (DSB) sites. The synaptonemal complex (SC) forms between paired chromosomes at zygonema, and full synapsis is reached at pachynema. TOPOISOMERASE II (TOPII) activity is essential for removal of the interlocks formed when homologous chromosomes trap other chromosomes in between them. During late prophase I (diplonema/diakinesis), the SC disappears and further condensing homologous chromosomes are held together by chiasmata. During anaphase I, loss of cohesion between the arms of sister chromatids allows the segregation of homologous chromosomes to the opposite poles. Centromeric cohesion is released at the onset of anaphase II, and sister chromatids segregate to form a tetrad.

early prophase I. Clustering of telomeres around the nucleolus allows pairing at the same time as when axial elements of the SC are assembled (Roberts *et al.*, 2009). However, in other species, the subtelomeric regions undergo differential behavior during pre-meiotic $G_2$ and prophase I (Colas *et al.*, 2008; Richards *et al.*, 2012). In the large genome of cereals, the telomere bouquet precedes chromosomes synapsis (Phillips *et al.*, 2012; Barakate *et al.*, 2014) and, although it is not required for pairing of homologous chromosomes, it may facilitate this process (Golubovskaya *et al.*, 2002). In this context, HR and synapsis start in the distal regions of the chromosomes in barley, but it has been suggested that this is likely to be related to the heterochromatin/euchromatin replication program rather than the telomere movements (Higgins *et al.*, 2012).

SMC complexes are essential during meiosis. Both condensin I and II complexes are important for maintaining the structure of meiotic chromosomes. Condensin I ensures normal condensation in centromeric and 45S rDNA regions, whereas condensin II eliminates interchromosome connections (Smith *et al.*, 2014). In addition, the cohesin complex is indispensable for proper pairing and HR (Golubovskaya *et al.*, 2006). Several meiosis-specific cohesin proteins have been identified in plants (Bolaños–Villegas *et al.*, 2017), but it is unknown how the replacement of the respective mitotic proteins takes place. ABSENCE OF FIRST MEIOTIC DIVISION 1 (AFD1), the meiosis-specific maize kleisin protein, is required for elongation of axial elements of the synaptonemal complex and also for normal bouquet formation (Golubovskaya *et al.*, 2006). In rice, if centromere cohesion is compromised, chromatids separate prematurely at anaphase I and chromosomes are intertwined, leading to chromosome bridges and fragmentation (Shao *et al.*, 2011). Mutants deficient for Arabidopsis SYNAPTIC 1 (SYN1), a meiosis-specific α-kleisin, present defects in arm cohesion during prophase I and problems in centromere cohesion from anaphase I onwards (Bai *et al.*, 1999; Cai *et al.*, 2003). In order to protect premature SYN1 depletion and thus cohesion at centromeres, SYN1 needs to be dephosphorylated by the protein phosphatases PP2AB'α and PP2AB'β (Yuan *et al.*, 2018). Precocious separation of sister chromatids at centromeres is also avoided by SHUGOSHIN–LIKE 1 and 2 (SGOL1 and SGOL2), and PATRONUS 1 (PANS1) (Cromer *et al.*, 2013; Zamariola *et al.*, 2014). This function is most probably conserved in both mitosis and meiosis, as shown in rice (Wang *et al.*, 2011). In Arabidopsis, absence of functional ESTABLISHMENT OF COHESION 1/CHROMOSOME TRANSMISSION FIDELITY 7 (ECO1/CTF7), involved in the establishment of chromatid cohesion, also produces a severe reduction of cohesion during meiosis (Bolaños–Villegas *et al.*, 2013). Furthermore, mutations in the two Arabidopsis *WAPL* genes, with a significant role in the removal of cohesin, lead to alterations in the organization of heterochromatin and delayed cohesin removal during prophase I (De *et al.*, 2014). Concerning the SMC5/6 complex, the SUMO (Small Ubiquitin-like Modifier) E3 ligase activity conferred by METHYL METHANE SULFONATE SENSITIVITY 21 (MMS21) and NSE4A kleisin is required for normal meiotic progression and gametophyte development in Arabidopsis (Liu *et al.*, 2014; Díaz *et al.*, 2019; Zelkowski *et al.*, 2019).

Most of the information on the behavior of chromatin in meiosis derives from studies with fixed cells. However, innovative methodologies are being developed to enable the dynamic analysis of meiotic processes in live meiocytes. In a pioneer study, prophase I has been analyzed within PMCs of intact anthers in maize (Sheehan and Pawlowski, 2009) and recently live microscopy of male meiosis was performed at high resolution in Arabidopsis (Prusicki *et al.*, 2019). Such advancements in technology will allow an in-depth analysis of the dynamics of meiotic processes. Finally, the link between chromatin conformation and gene regulation during meiosis is still very obscure despite the number of genomic and transcriptomic studies in various plant species (Zhou and Pawlowski, 2014). However, most of these analyses have mainly been done with tissue covering the overall meiosis rather than specific meiotic stages, which is necessary to understand the gene expression pattern. In addition, transcriptomic studies would also benefit from complementary proteomic experiments to address the regulation of gene/protein meiotic networks.

## Chromatin dynamics during reproductive development

In Angiosperms, sexual reproduction starts with the development of flowers, when the SAM is transformed into the inflorescence meristem (IM) continuously producing the floral meristems (FMs). Remarkably, the FM switches from an indeterminate fate to a determinate fate to give rise to all the organs of the flower, the gametes, and the fruit. All reproductive development transitions are controlled by endogenous, hormonal, or external environmental signaling pathways, which require complex gene regulatory networks involving transcription factors and epigenetic mechanisms.

The floral initiation is precisely coordinated via a complex gene network that integrates the age, photoperiod, temperature, and hormonal signals (Andrés and Coupland, 2012). Under favorable conditions, the Arabidopsis systemic floral activator *FLOWERING LOCUS T* (*FT*; the florigen) or its orthologs in other species (e.g. *VRN3* in cereals) change SAMs to IMs. In Arabidopsis, FT expression is subjected to photoperiod and ambient temperature, and is under a complex balance of active and repressive chromatin modifications involving both Polycomb Repressive Complex (PRC) 1 and 2 (He, 2012). Expression of the FT target and flowering pathway integrator, SUPPRESSION OF OVEREXPRESSION OF CONSTANS 1 (SOC1), turns on the FM identity genes *APETALA 1* (*AP1*) and *LEAFY*, which promote the formation of the floral primordium (reviewed in Guo *et al.*, 2015). The homeodomain transcription factor WUSCHEL (WUS) plays a central role in the process of FM determinacy by specifying the maintenance of stem cell activity within the organizing center of the SAM, IM, and FM (Cao *et al.*, 2015). In cooperation with LEAFY, WUS activates the MADS-box transcription factor gene *AGAMOUS* (*AG*), which initiates the reproductive organ development. Thereafter, AG represses *WUS* activity to ensure termination of the FM, and to promote all the finely tuned developmental transitions required

for the proper development of floral organs. The repression of *WUS* is a perfect example to illustrate the importance of epigenetic regulatory mechanisms during FM termination. First, AG binds to the *WUS* locus, which allows the recruitment of the PRC2 catalytic subunit CURLY LEAF to mediate the deposition of H3K27me3 repressive marks on *WUS*. Then components of the PRC1 complex recognize H3K27me3, which results in the compaction of chromatin and further *WUS* repression. Thereafter, AG turns on the C2H2 zinc-finger transcription factor *KNUCKLES* gene (*KNU*), which terminates the inflorescence by stabilizing WUS repression (Bollier *et al.*, 2018).

After meiosis (see the previous section), the male haploid gametophyte (microspore) undergoes an asymmetric division to produce a generative cell (GC) and a vegetative cell (VC), and the GC divides once more to produce two sperm cells (SCs) representing the male gametes (reviewed, for example, by Berger and Twell, 2011). SCs and VCs have very different chromatin characteristics, which also determine their fate, genome integrity, and capacity to divide (Slotkin *et al.*, 2009; Calarco *et al.*, 2012; Ibarra *et al.*, 2012). The SC nuclei are very compact and strongly repress transposons by maintaining high levels of H3K9me2, and CG and CHG methylation (Schoft *et al.*, 2009; Calarco *et al.*, 2012; Ibarra *et al.*, 2012; Hsieh *et al.*, 2016), whereas CHH methylation is generally low, but shows complex dynamics with temporal increases (Walker *et al.*, 2018). In contrast, the VC nuclei are de-condensed, without CenH3, H3K9me2, and DECREASED IN DNA METHYLATION 1 (DDM1), but rich in 21 nt siRNAs, suggesting loss of competence to divide, strongly reduced maintenance methylation control, and activation of the non-canonical RdDM pathway (Schoft *et al.*, 2009; Slotkin *et al.*, 2009; Creasey *et al.*, 2014). This leads to decreased CG methylation and increased CHH methylation levels and transcriptional activation of transposable elements (TEs) in VCs (Mosher *et al.*, 2009; Slotkin *et al.*, 2009; Calarco *et al.*, 2012; Creasey *et al.*, 2014; Martínez *et al.*, 2016, 2018). Furthermore, VCs show enrichment in H3K27me3, indicating high PRC2 activity (Borg and Berger, 2015). The functional significance of such extensive epigenetic reprogramming is still debated, but the activation of TEs in VCs may represent a non-autonomous silencing mechanism, which switches off any potentially active transposons in the germline and thus preserves the genome integrity of the next generation. However, to what extent this is typical for plants other than Arabidopsis remains unknown. For example, cereals lack specific epigenetic factors present in Arabidopsis such as DEMETER (DME) or CHROMOMETHYLASE 2 (CMT2), but have multiple copies of other factors including DNA METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLASE 3 (CMT3), DDM1, or specific subunits of Pol IV and Pol V (Zemach *et al.*, 2010, 2013; Li *et al.*, 2014; Haag *et al.*, 2014; Shi *et al.*, 2014; Bewick and Schmitz, 2017). In addition, the same factors in cereals may have different effects on DNA methylation, such as ZmDDM1 that is required for the formation of mCHH islands via the RdDM pathway (Fu *et al.*, 2018; Long *et al.*, 2019). All this indicates a diversification and/or specialization of functions and a more important role for the small RNAs in epigenetic

programming of cereal pollen. In rice SCs, there is high expression from *OsDRM2* and a new small RNA pathway involving a non-canonical ARGONAUTE (AGO) and DICER-LIKE (DCL3) proteins, suggesting high CHH methylation levels (Russell *et al.*, 2012; Anderson *et al.*, 2013). In addition, there seem to be a specific variant of the largest subunit of Pol V in grasses (Trujillo *et al.*, 2018), and future studies will reveal whether these factors act in a novel RdDM pathway. Long intergenic non-coding RNAs (lincRNAs), correlated with H3K27me3, have been identified in the rice male gametophyte (Zhang *et al.*, 2014; Johnson *et al.*, 2018). The high and medium numbers of copies of putative orthologs of H3K27 and H3K4 demethylases, respectively, indicates that rice SCs may require more extensive reprogramming of repressive marks (Anderson *et al.*, 2013).

The replacement of canonical histones by specific variants is also characteristic of epigenetic control at male gametogenesis. In Arabidopsis SCs, the histone H3 variant, MALE GAMETE-SPECIFIC HISTONE 3 (MGH3), is the most abundant (Okada *et al.*, 2005; Ingouff *et al.*, 2007; Ingouff and Berger, 2010). This variant has been correlated with the loss of H3K27me3 methylation, due to the composition of the adjacent amino acid residues (Borg and Berger, 2015). In rice, a specific combination of H2A, H2B, and H3 histone proteins has also been identified in SCs (Russell *et al.*, 2012; Anderson *et al.*, 2013). Histones H3.709 and H2A.Z are the most remarkable. Histone H3.709, although probably an ortholog of MGH3, is quite divergent in its amino acid composition. Replacement of histones also occurs in the Arabidopsis VC, since CenH3 is progressively lost in centromeric heterochromatin when it begins to de-condense, while there is a loss of H3K9me2 marks, indicating a state of terminal differentiation (Schoft *et al.*, 2009; Mérai *et al.*, 2014). However, CenH3 and H3K9me2 persist in VCs of rye and barley (Houben *et al.*, 2011; Pandey *et al.*, 2013), probably reflecting a temporal shift between pollination and fertilization in these species (Borg and Berger, 2015). In maize, the haploid microspores carrying a knockdown mutation in *hda108* gene collapsed and failed to develop properly, indicating that histone acetylation/deacetylation affects microspore viability (Forestan *et al.*, 2018). In *Brassica rapa*, H3K4me3 and H3K27me3 deposition is necessary for the regulation of the pollen wall construction (Shen *et al.*, 2019).

The female gametophyte develops in the ovule according to the Polygonum type in ~70% of flowering plants, including, for example, Arabidopsis, maize, rice, wheat, and soybean. In megasporogenesis, the diploid megaspore mother cell undergoes meiosis, resulting in four haploid megaspores. One megaspore develops into the female gametophyte, while the others die. The formation and differentiation of the different cell types in the reproductive lineage are characterized by global changes in chromatin organization. Histone modifications were observed via cytogenetic and chromatin reporter studies in Arabidopsis megaspores and also in the surrounding nucellar cells in maize (Garcia-Aguilar *et al.*, 2010; She *et al.*, 2013). Genetic analyses have identified DNA methylation acting upon establishment of the megaspore fate, and also the action of small RNAs silencing TEs in the female gametes in Arabidopsis and maize (Garcia-Aguilar *et al.*, 2010;

Olmedo-Monfil *et al.*, 2010). The multicellular embryo sac consists of the egg cell, the central cell, two synergid cells, and three antipodal cells. The female gametes exhibit chromatin dimorphism as they express different histone H3 proteins, with the egg cell expressing only the H3.3 variant, whereas there are both H3.1 and H3.3 proteins in the central cell (Ingouff and Berger, 2010). Due to the technically limiting accessibility to the female gametophyte, gene-level resolution of the chromatin perturbations has not been reported to date. The histone modifications observed suggest a global epigenetic reprogramming phase during development of the female gametophyte. The epigenetic dimorphism of the two female gametes at the DNA methylation level, with the global demethylation of the central cell versus the non-CG DNA methylation of the egg cell, highlights the different roles which these two cell types are going to play in seed development (Pillot *et al.*, 2010). For an extensive review on the dynamics of the chromatin landscape on the female gametophyte development follow Baroux and Autran (2015).

In the zygote, the parentally derived histone H3 variants are replaced before the first division of the embryo to reflect the content found in sporophytic cells (Ingouff and Berger, 2010). Two maternal epigenetic pathways are acting in the early embryo to regulate the paternal transcripts, the RdDM pathway and the histone chaperone complex chromatin assembly factor 1 (CAF1). These pathways do not regulate genomic imprinting (Autran *et al.*, 2011). The central cell will give rise (upon fusion with one sperm cell nucleus) to the endosperm. In the endosperm, maternally expressed genes will be suppressed by the PRC2 complex, including the central cell lineage-specific H3K27 methyltransferase FERTILIZATION INDEPENDENT SEED 1/MEDEA (FIS1/MEA), implicated in the regulation of type I MADS-box genes and transition from the syncytial to cellularized stage (S, Zhang *et al.*, 2018). It should be noted that endosperm development is sensitive to parental genome dosage, and the majority of imprinted genes reported are expressed from the maternal genome in the endosperm (reviewed, for example, in Gehring and Satyaki, 2017). Endosperm chromatin is characterized by a looser structure, DNA hypomethylation, and decreased levels of H3K9me2, when compared with somatic tissues and embryo (Baroux *et al.*, 2007; Pillot *et al.*, 2010). In contrast to embryo development, extensive demethylation occurs during endosperm development and this dynamic process allows for imprinting variation observed in maize and Arabidopsis (Gehring *et al.*, 2009; Waters *et al.*, 2013; Pignatta *et al.*, 2018). In maize, HDA101 and members of different chromatin-remodeling complexes affect endosperm transfer cells leading to an alteration in the kernel size (Yang *et al.*, 2016). Kernels of *hda108 hda101* plants showed a strong defective phenotype with fully or partially empty pericarp. Starchy endosperm tissue failed to accumulate starch or developed only partially in defective kernels, while the embryo showed abnormalities that varied from the presence of an undifferentiated aborted embryo to a defective embryo blocked at the coleoptilar stage (Forestan *et al.*, 2018).

Seeds are embedded in fruits, many of which are an important source of food for humans. The best understood development of fleshy fruits is that of tomato, which displays remarkable characteristics related to chromosome structure, chromatin organization, and chromatin dynamics (Bourdon *et al.*, 2012). A major developmental feature is an increase in nuclear DNA content due to endoreduplication leading to cell hypertrophy, thereby influencing fruit growth and size (Chevalier *et al.*, 2014). Whether chromatin modifications are associated with endoreduplication still remains largely unknown. However, it was shown in Arabidopsis that endoreduplicated nuclei have less condensed heterochromatin (Schubert *et al.*, 2006; Jégu *et al.*, 2013). In tomato, DNA methylation decreases in the highly endoreduplicated pericarp tissue and is significantly reduced at the onset of fruit maturation and during ripening (Teyssier *et al.*, 2008; Zhong *et al.*, 2013), possibly to control the gene expression according to a tissue-specific endoreduplication status. Ectopic overexpression of the DAMAGED DNA BINDING PROTEIN 1 (DDB1), a member of the DDB1–CUL4-based E3 ubiquitin ligase complex, regulating many developmental processes via chromatin remodeling, decreased the size of flowers and fruits in tomato (Liu *et al.*, 2012) via up-regulation of two positive regulators of endoreduplication *SlWEE1* and *SlCCS52A* (Azzi *et al.*, 2015). Currently, there is increasing evidence for epigenetic control during fruit organogenesis, and epigenome dynamics play an important role during fruit maturation and ripening in tomato (reviewed in Giovannoni *et al.*, 2017).

## Plant chromatin modifications for the purposes of plant breeding

Decades of breeding and selection have narrowed down the pool of genetic variability in many crops (Palmgren *et al.*, 2015). Crop breeding programs have classically relied on sequence-based genetic variability of either natural or induced origin. These efforts have allowed the generation of varieties with an increased and more stable yield, and relatively well adapted to biotic and abiotic stresses. However, the exploitation of genetic variability existing within gene pools has been limited. Furthermore, not all the heritable phenotypic diversity can be explained by sequence variation, and has been termed the missing heritability (Maher, 2008; Gallusci *et al.*, 2017). Such variation could have an epigenetic basis.

The applicability of chromatin modifications for the purpose of crop improvement (Fig. 3) depends on their stability and heritability as the two key features. Epigenetic modifications may be of interest for breeders only if their regulatory effects are maintained through mitosis and ideally through meiosis. Here, DNA methylation and specific histone PTMs are the prime candidates for crop improvement, as they were mitotically transmittable for at least a limited time in several species (Hyun *et al.*, 2013; Gaydos *et al.*, 2014; Avramova, 2015; Jiang and Berger, 2017; Kawakatsu *et al.*, 2017). This raises the possibility of employing them as tools for breeding in clonally propagated crops, such as many fruit trees. However, for seed-propagated crops, specific chromatin modifications need to pass the epigenetic resetting barriers during gametogenesis and seed development in order to pass to the next generation
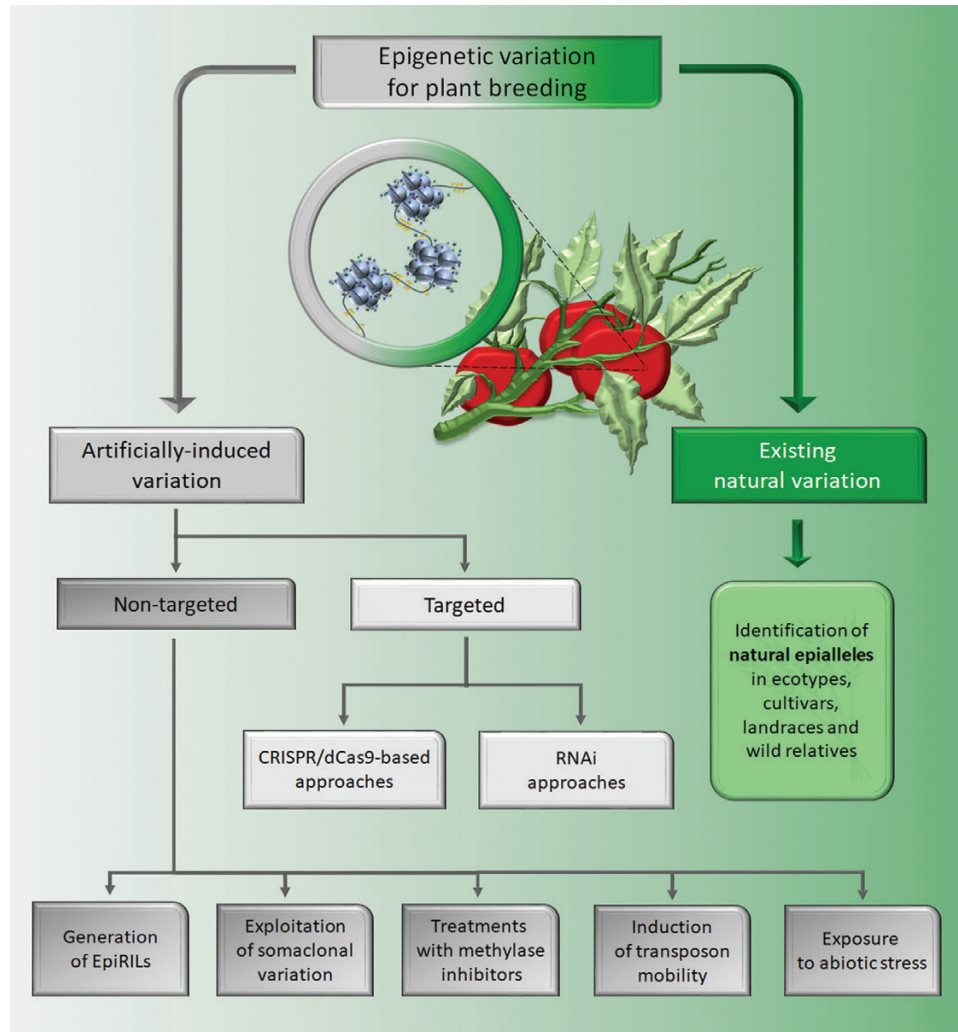
**Fig. 3.** Applications of epigenetic variation for the purposes of plant breeding. Natural epigenetic variation is relatively little explored and known cases were often selected by the phenotype and only later described to have an epigenetic basis. Presumably, genome-wide screening for natural epigenetic variation will allow less biased use of the naturally occurring germplasms in the future. In contrast, induced epigenetic variation is provoked by humans either in a targeted manner towards a specific genomic locus or in an untargeted manner with subsequent identification and selection of the modified loci. Choice of the method(s) is guided by the purpose, the species, and its available resources. Some of the artificially produced epialleles fall under the GMO regulations.

(Pecinka and Mittelsten Scheid, 2012; Grossniklaus *et al.*, 2013; Iwasaki and Paszkowski, 2014; Kawashima and Berger, 2014; Quadrana and Colot, 2016; Roessler *et al.*, 2018). Here, DNA methylation seems to be the best candidate due to its stability and because PTMs are lost due to gametogenesis specific-removal and replacement of the parental nucleosomes (Ingouff *et al.*, 2010; Quadrana and Colot, 2016).

Plant developmental processes determine a great number of traits of agronomic interest that have been targeted for selection in crops. Some of them are epigenetically regulated, either by DNA methylation or histone PTMs such as leaf shape, flowering time and flower development, male fertility, oil yield, fruit ripening, grain size, plant stature, inflorescence structure, branching plant architecture, boll setting rate, abscission rate, photoperiod responses, etc. (Zhang, 2012; Ong-Abdullah *et al.*, 2015; Xianwei *et al.*, 2015; Bull *et al.*, 2017; Latrasse *et al.*, 2017b; van Esse *et al.*, 2017; Fan *et al.*, 2018; Song *et al.*, 2018). Expanding evidence also shows that epigenetic control has an important role in the fine-tuning of the responses to

biotic and abiotic stress (Gourcilleau *et al.*, 2010; Kim *et al.*, 2010; González *et al.*, 2011, 2013; Ding and Wang, 2015). This raises the possibility of generating or selecting variability of epigenetic changes to assist plant breeding. Stably inherited epialleles have been characterized for genes controlling some developmental processes. Examples of such epialleles in crops include: the tomato *CNR* locus controlling fruit ripening (Manning *et al.*, 2006); oil palm *MANTLED* that regulates oil yield (Ong-Abdullah *et al.*, 2015); cotton *CONSTANS-LIKE 2* that determines photoperiod sensibility (Song *et al.*, 2017); rice *FERTLIZATION INDEPENDENT ENDOSPERM 1* (*FIE1*), which regulates plant height and flower development (Zhang *et al.*, 2012); *RAV6* affecting leaf angle and grain size (Xianwei *et al.*, 2015); or *SEMI-ROLLED LEAF 1* (*SRL1*), which determines rice cell wall formation (Li *et al.*, 2017).

Thus, epigenetic modifications are a source of phenotypic diversity and it is desirable to identify and/or generate novel epialleles of interest for crop improvement (Fig. 3). One possible approach is to select epigenetic variants among the

natural diversity by exploiting DNA methylation states in different germplasms (Takuno *et al.*, 2016). This type of analysis has revealed large amounts of epigenetic variability in ecotypes, cultivars, landraces, and wild relatives (Eichten *et al.*, 2013; Schmitz *et al.*, 2013b; Garg *et al.*, 2015; Venetsky *et al.*, 2015; Kumar *et al.*, 2017; Song *et al.*, 2017; Liu *et al.*, 2018; Shen *et al.*, 2018). However, it requires good reference genomes and can be more time-consuming and tedious than mining genetic polymorphisms. The easiest way to link DNA methylation polymorphisms with phenotypes is to simultaneously monitor gene expression (Eichten *et al.*, 2013; Song *et al.*, 2017). However, this may be challenging for genes with tissue-specific transcription.

Epialleles can also be generated artificially. Untargeted approaches employ cell culture (Mittelsten Scheid *et al.*, 2003; Ong-Abdullah *et al.*, 2015; Li *et al.*, 2017; Coronel *et al.*, 2018), abiotic stresses (Verkest *et al.*, 2015), transposon mobilization (Thieme *et al.*, 2017), or treatment with specific epigenetic inhibitors (Marfil *et al.*, 2012; Baubec *et al.*, 2014; Pecinka and Liu, 2014; Xu *et al.*, 2016; Zhu *et al.*, 2018). In addition, this can be achieved by the generation of epigenetic recombinant inbred lines (epiRILs) from crosses between the wild type and maintenance DNA methylation mutants. Although epiRILs are a well-established system in Arabidopsis (Dapp *et al.*, 2015; Zhang *et al.*, 2016; Lauss *et al.*, 2018; Zhang *et al.*, 2018), their use in crops is still in its infancy and might be influenced by the reproductive modality (Schmitz *et al.*, 2013a) and availability of viable epiregulator mutants (Anderson *et al.*, 2018). However, the current trends are directed towards controlled induction of the chromatin states. RNAi allows directing DNA methylation to specific positions and thus silencing the target loci. In addition, there are studies demonstrating that the modified CRISPR (clustered regularly interspaced short palindromic repeat) system using Cas9 or related nucleases (such as Cfp1) offers wide possibilities to change chromatin at specific loci (Liu and Moschou, 2018; Xie *et al.*, 2018). In this approach, chromatin remodelers, DNA or histone (de)methylases, transcription factors, or specific protein domains can be, directly or via a marker peptide–antibody-based system, fused to the catalytically dead Cas9 (dCas9), which leads to the recruitment of dCas9 to the locus of interest and chromatin change (Gallego-Bartolomé *et al.*, 2018; Liu and Moschou, 2018; Xie *et al.*, 2018). We predict that the number of dCas9-induced modifications will grow rapidly in the model plants as well as in crops. This approach has a great potential to shed more light on how the chromatin states are established, maintained, and erased in plants. In addition, this could improve agriculturally relevant developmental or stress resistance-related traits in crops; however, the legal restrictions will most probably remain the main hurdle towards practical use of such inventions world-wide.

Chromatin modifications have emerged as a complementary source of variability contributing to plant phenotypic plasticity (Fig. 3). It could also address new challenges in crop improvement, including adaptive responses to environmental stresses. Since the emergence and inheritance of epigenetic variation differs from the genetic variants, current methods of trait mapping miss substantial phenotype-determining variation and thus may have reduced efficacy. Therefore, the relative contribution of genetic versus epigenetic variation remains unknown (Pecinka *et al.*, 2013). However, plant breeding using chromatin traits can be assisted by newly developed tools including process-based models (Hu *et al.*, 2015; Gallusci *et al.*, 2017), or epigenome-wide association studies (EWAS) (Rakyan *et al.*, 2011).

# Future perspectives in plant breeding strategies

Classical plant breeding harnesses the genetic variation that is generated by homologous recombination during meiosis. For example, in cereals, a high amount of 20–30% (according to some sources up to 50%) of genes rarely recombine (Sandhu and Gill, 2002; International Barley Genome Sequencing Consortium, 2012; Higgins *et al.*, 2014; Mascher *et al.*, 2017), limiting the genetic diversity available for plant breeders and breaking the desirable combination of alleles in elite cultivars (Mascher *et al.*, 2017; Appels *et al.*, 2018; Ramírez-González *et al.*, 2018). In this context, a better understanding about the influence of the epigenetic make up on meiotic recombination would contribute to development of novel strategies to modify the recombination pattern and to generate new elite crop varieties (Fig. 3). The ever-increasing knowledge drawn from epigenetics studies in model and crop plants paves the way to applied perspectives and foreseen plant breeding strategies. The exploitation of epigenetic diversity is the forthcoming challenge for the next plant breeding strategies, since chromatin modifications are tightly intertwined with plant phenotypic plasticity (reviewed in Pecinka *et al.*, 2013; Gallusci *et al.*, 2017). To cope with the improvement of genetic diversity resulting from intense plant breeding programs, epigenetic diversity may thus provide this opportunity to select for new traits related to plant adaptation to environmental constraints, crop yield, or quality of plant products, pending a better understanding of all the associated regulatory mechanisms.

## Acknowledgements

# References

**Agashe B, Prasad CK, Siddiqi I.** 2002. Identification and analysis of DYAD: a gene required for meiotic chromosome organisation and female meiotic progression in Arabidopsis. Development **129**, 3935–3943.

**Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, Sundaresan V.** 2013. Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. The Plant Journal **76**, 729–741.

**Anderson SN, Zynda GJ, Song J, Han Z, Vaughn MW, Li Q, Springer NM.** 2018. Subtle perturbations of the maize methylome reveal genes and transposons silenced by chromomethylase or RNA-directed DNA methylation pathways. G3 **8**, 1921–1932.

**Andrés F, Coupland G.** 2012. The genetic basis of flowering responses to seasonal cues. Nature Reviews. Genetics **13**, 627–639.

**Andreuzza S, Nishal B, Singh A, Siddiqi I.** 2015. The chromatin protein DUET/MMD1 controls expression of the meiotic gene TDM1 during male meiosis in Arabidopsis. PLoS Genetics **11**, e1005396.

**Appels R, Eversole K, Feuillet C, et al.** 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science **361**, eaar7191.

**Armstrong SJ, Franklin FC, Jones GH.** 2001. Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*. Journal of Cell Science **114**, 4207–4217.

**Autran D, Baroux C, Raissig MT, et al.** 2011. Maternal epigenetic pathways control parental contributions to Arabidopsis early embryogenesis. Cell **145**, 707–719.

**Avramova Z.** 2015. Transcriptional 'memory' of a stress: transient chromatin and memory (epigenetic) marks at stress-response genes. The Plant Journal **83**, 149–159.

**Azzi L, Deluche C, Gévaudant F, Frangne N, Delmas F, Hernould M, Chevalier C.** 2015. Fruit growth-related genes in tomato. Journal of Experimental Botany **66**, 1075–1086.

**Bai X, Peirson BN, Dong F, Xue C, Makaroff CA.** 1999. Isolation and characterization of SYN1, a RAD21-like gene essential for meiosis in Arabidopsis. The Plant Cell **11**, 417–430.

**Baker K, Dhillon T, Colas I, Cook N, Milne I, Milne L, Bayer M, Flavell AJ.** 2015. Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. The Plant Journal **84**, 111–124.

**Barakate A, Higgins JD, Vivera S, et al.** 2014. The synaptonemal complex protein ZYP1 is required for imposition of meiotic crossovers in barley. The Plant Cell **26**, 729–740.

**Baroux C, Autran D.** 2015. Chromatin dynamics during cellular differentiation in the female reproductive lineage of flowering plants. The Plant Journal **83**, 160–176.

**Baroux C, Pien S, Grossniklaus U.** 2007. Chromatin modification and remodeling during early seed development. Current Opinion in Genetics & Development **17**, 473–479.

**Bass HW, Riera-Lizarazu O, Ananiev EV, Bordoli SJ, Rines HW, Phillips RL, Sedat JW, Agard DA, Cande WZ.** 2000. Evidence for the coincident initiation of homolog pairing and synapsis during the telomere-clustering (bouquet) stage of meiotic prophase. Journal of Cell Science **113 (Pt 6)**, 1033–1042.

**Baubec T, Finke A, Mittelsten Scheid O, Pecinka A.** 2014. Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. EMBO Reports **15**, 446–452.

**Berger F, Twell D.** 2011. Germline specification and function in plants. Annual Review of Plant Biology **62**, 461–484.

**Bewick AJ, Schmitz RJ.** 2017. Gene body DNA methylation in plants. Current Opinion in Plant Biology **36**, 103–110.

**Boateng KA, Yang X, Dong F, Owen HA, Makaroff CA.** 2008. SWI1 is required for meiotic chromosome remodeling events. Molecular Plant **1**, 620–633.

**Boden SA, Kavanová M, Finnegan EJ, Wigge PA.** 2013. Thermal stress effects on grain yield in *Brachypodium distachyon* occur via H2A.Z-nucleosomes. Genome Biology **14**, R65.

**Bolaños-Villegas P, De K, Pradillo M, Liu D, Makaroff CA.** 2017. In favor of establishment: regulation of chromatid cohesion in plants. Frontiers in Plant Science **8**, 846.

**Bolaños-Villegas P, Yang X, Wang HJ, Juan CT, Chuang MH, Makaroff CA, Jauh GY.** 2013. Arabidopsis CHROMOSOME TRANSMISSION FIDELITY 7 (AtCTF7/ECO1) is required for DNA repair, mitosis and meiosis. The Plant Journal **75**, 927–940.

**Bollier N, Sicard A, Leblond J, et al.** 2018. At-MINI ZINC FINGER2 and Sl-INHIBITOR OF MERISTEM ACTIVITY, a conserved missing link in the regulation of floral meristem termination in Arabidopsis and tomato. The Plant Cell **30**, 83–100.

**Borg M, Berger F.** 2015. Chromatin remodelling during male gametophyte development. The Plant Journal **83**, 177–188.

**Bourbousse C, Mestiri I, Zabulon G, Bourge M, Formiggini F, Koini MA, Brown SC, Fransz P, Bowler C, Barneche F.** 2015. Light signaling controls nuclear architecture reorganization during seedling establishment. Proceedings of the National Academy of Sciences, USA **112**, E2836–E2844.

**Bourdon M, Pirrello J, Cheniclet C, et al.** 2012. Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. Development **139**, 3817–3826.

**Bull H, Casao MC, Zwirek M, et al.** 2017. Barley SIX-ROWED SPIKE3 encodes a putative Jumonji C-type H3K9me2/me3 demethylase that represses lateral spikelet fertility. Nature Communications **8**, 936.

**Cai X, Dong F, Edelmann RE, Makaroff CA.** 2003. The Arabidopsis SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing. Journal of Cell Science **116**, 2999–3007.

**Calarco JP, Borges F, Donoghue MT, et al.** 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell **151**, 194–205.

**Cao X, He Z, Guo L, Liu X.** 2015. Epigenetic mechanisms are critical for the regulation of WUSCHEL expression in floral meristems. Plant Physiology **168**, 1189–1196.

**Casati P, Campi M, Chu F, Suzuki N, Maltby D, Guan S, Burlingame AL, Walbot V.** 2008. Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. The Plant Cell **20**, 827–842.

**Casati P, Stapleton AE, Blum JE, Walbot V.** 2006. Genome-wide analysis of high-altitude maize and gene knockdown stocks implicates chromatin remodeling proteins in response to UV-B. The Plant Journal **46**, 613–627.

**Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O.** 2014. How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genetics **10**, e1004115.

**Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N.** 2014. Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. Journal of Experimental Botany **65**, 2731–2746.

**Choi K, Zhao X, Kelly KA, et al.** 2013. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nature Genetics **45**, 1327–1336.

**Choi K, Zhao X, Tock AJ, et al.** 2018. Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. Genome Research **28**, 532–546.

**Choi SM, Song HR, Han SK, Han M, Kim CY, Park J, Lee YH, Jeon JS, Noh YS, Noh B.** 2012. HDA19 is required for the repression of salicylic acid biosynthesis and salicylic acid-mediated defense responses in Arabidopsis. The Plant Journal **71**, 135–146.

**Colas I, Shaw P, Prieto P, Wanous M, Spielmeyer W, Mago R, Moore G.** 2008. Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. Proceedings of the National Academy of Sciences, USA **105**, 6075–6080.

**Coleman-Derr D, Zilberman D.** 2012. Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. PLoS Genetics **8**, e1002988.

**Coronel CJ, González AI, Ruiz ML, Polanco C.** 2018. Analysis of somaclonal variation in transgenic and regenerated plants of *Arabidopsis thaliana* using methylation related metAFLP and TMD markers. Plant Cell Reports **37**, 137–152.

**Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA.** 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature **508**, 411–415.

**Cromer L, Jolivet S, Horlow C, Chelysheva L, Heyman J, De Jaeger G, Koncz C, De Veylder L, Mercier R.** 2013. Centromeric cohesion is protected twice at meiosis, by SHUGOSHINs at anaphase I and by PATRONUS at interkinesis. Current Biology **23**, 2090–2099.

**Dapp M, Reinders J, Bédiée A, Balsera C, Bucher E, Theiler G, Granier C, Paszkowski J.** 2015. Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids. Nature Plants **1**, 15092.

**De K, Bolaños-Villegas P, Mitra S, Yang X, Homan G, Jauh GY, Makaroff CA.** 2016. The opposing actions of Arabidopsis CHROMOSOME TRANSMISSION FIDELITY7 and WINGS APART-LIKE1 and 2 differ in mitotic and meiotic cells. The Plant Cell **28**, 521–536.

**De K, Sterle L, Krueger L, Yang X, Makaroff CA.** 2014. *Arabidopsis thaliana* WAPL is essential for the prophase removal of cohesin during meiosis. PLoS Genetics **10**, e1004497.

**Dhawan R, Luo H, Foerster AM, Abuqamar S, Du HN, Briggs SD, Mittelsten Scheid O, Mengiste T.** 2009. HISTONE MONOUBIQUITINATION1 interacts with a subunit of the mediator complex and regulates defense against necrotrophic fungal pathogens in Arabidopsis. The Plant Cell **21**, 1000–1019.

**Diallo AO, Ali-Benali MA, Badawi M, Houde M, Sarhan F.** 2012. Expression of vernalization responsive genes in wheat is associated with histone H3 trimethylation. Molecular Genetics and Genomics **287**, 575–590.

**Díaz M, Pečinková P, Nowicka A, Baroux C, Sakamoto T, Gandha PY, Jeřábková H, Matsunaga S, Grossniklaus U, Pecinka A.** 2019. SMC5/6 complex subunit NSE4A is involved in DNA damage repair and seed development in Arabidopsis. The Plant Cell **31**, 1579–1597.

**Ding B, Wang GL.** 2015. Chromatin versus pathogens: the function of epigenetics in plant immunity. Frontiers in Plant Science **6**, 675.

**Ding Y, Fromm M, Avramova Z.** 2012. Multiple exposures to drought 'train' transcriptional responses in Arabidopsis. Nature Communications **3**, 740.

**Dixon JR, Gorkin DU, Ren B.** 2016. Chromatin domains: the unit of chromosome organization. Molecular Cell **62**, 668–680.

**Doğan ES, Liu C.** 2018. Three-dimensional chromatin packing and positioning of plant genomes. Nature Plants **4**, 521–529.

**Dong Q, Li N, Li X, *et al*.** 2018. Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. The Plant Journal **94**, 1141–1156.

**Du J, Zhong X, Bernatavichute YV, *et al*.** 2012. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151**, 167–180.

**Dutta A, Choudhary P, Caruana J, Raina R.** 2017. JMJ27, an Arabidopsis H3K9 histone demethylase, modulates defense against *Pseudomonas syringae* and flowering time. The Plant Journal **91**, 1015–1028.

**Edmondson DG, Davie JK, Zhou J, Mirnikjoo B, Tatchell K, Dent SY.** 2002. Site-specific loss of acetylation upon phosphorylation of histone H3. Journal of Biological Chemistry **277**, 29496–29502.

**Eichten SR, Briskine R, Song J, *et al*.** 2013. Epigenetic and genetic influences on DNA methylation variation in maize populations. The Plant Cell **25**, 2783–2797.

**Eichten SR, Springer NM.** 2015. Minimal evidence for consistent changes in maize DNA methylation patterns following environmental stress. Frontiers in Plant Science **6**, 308.

**Fan S, Wang J, Lei C, Gao C, Yang Y, Li Y, An N, Zhang D, Han M.** 2018. Identification and characterization of histone modification gene family reveal their critical responses to flower induction in apple. BMC Plant Biology **18**, 173.

**Feitoza L, Costa L, Guerra M.** 2017. Condensation patterns of prophase/prometaphase chromosome are correlated with H4K5 histone acetylation and genomic DNA contents in plants. PLoS One **12**, e0183341.

**Feng CM, Qiu Y, Van Buskirk EK, Yang EJ, Chen M.** 2014. Light-regulated gene repositioning in Arabidopsis. Nature Communications **5**, 3027.

**Forestan C, Farinati S, Rouster J, Lassagne H, Lauria M, Dal Ferro N, Varotto S.** 2018. Control of maize vegetative and reproductive development, fertility, and rRNAs silencing by HISTONE DEACETYLASE 108. Genetics **208**, 1443–1466.

**Fransz P, De Jong JH, Lysak M, Castiglione MR, Schubert I.** 2002. Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. Proceedings of the National Academy of Sciences, USA **99**, 14584–14589.

**Friesner JD, Liu B, Culligan K, Britt AB.** 2005. Ionizing radiation-dependent gamma-H2AX focus formation requires ataxia telangiectasia mutated and ataxia telangiectasia mutated and Rad3-related. Molecular Biology of the Cell **16**, 2566–2576.

**Fu FF, Dawe RK, Gent JI.** 2018. Loss of RNA-directed DNA methylation in maize chromomethylase and DDM1-type nucleosome remodeler mutants. The Plant Cell **30**, 1617–1627.

**Fu Y, Luo GZ, Chen K, *et al*.** 2015. N$^6$-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. Cell **161**, 879–892.

**Gallego-Bartolomé J, Gardiner J, Liu W, Papikian A, Ghoshal B, Kuo HY, Zhao JM, Segal DJ, Jacobsen SE.** 2018. Targeted DNA demethylation of the Arabidopsis genome using the human TET1 catalytic domain. Proceedings of the National Academy of Sciences, USA **115**, E2125–E2134.

**Gallusci P, Dai Z, Génard M, Gauffretau A, Leblanc-Fournier N, Richard-Molard C, Vile D, Brunel-Muguet S.** 2017. Epigenetics for plant improvement: current knowledge and modeling avenues. Trends in Plant Science **22**, 610–623.

**Ganguly DR, Crisp PA, Eichten SR, Pogson BJ.** 2017. The Arabidopsis DNA methylome is stable under transgenerational drought stress. Plant Physiology **175**, 1893–1912.

**Garcia-Aguilar M, Michaud C, Leblanc O, Grimanelli D.** 2010. Inactivation of a DNA methylation pathway in maize reproductive organs results in apomixis-like phenotypes. The Plant Cell **22**, 3249–3267.

**Garg R, Narayana Chevala V, Shankar R, Jain M.** 2015. Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. Scientific Reports **5**, 14922.

**Gaydos LJ, Wang W, Strome S.** 2014. Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. Science **345**, 1515–1518.

**Gehring M, Bubb KL, Henikoff S.** 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science **324**, 1447–1451.

**Gehring M, Satyaki PR.** 2017. Endosperm and imprinting, inextricably linked. Plant Physiology **173**, 143–154.

**Gelei J.** 1921. Weitere Studien uber die Oogenese des Dendrocoelum lactum. Archiv fur Zellforschung **16**, 88–169.

**Giovannoni J, Nguyen C, Ampofo B, Zhong S, Fei Z.** 2017. The epigenome and transcriptional dynamics of fruit ripening. Annual Review of Plant Biology **68**, 61–84.

**Glazebrook J.** 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annual Review of Phytopathology **43**, 205–227.

**Golubovskaya IN, Hamant O, Timofejeva L, Wang CJ, Braun D, Meeley R, Cande WZ.** 2006. Alleles of afd1 dissect REC8 functions during meiotic prophase I. Journal of Cell Science **119**, 3306–3315.

**Golubovskaya IN, Harper LC, Pawlowski WP, Schichnes D, Cande WZ.** 2002. The pam1 gene is required for meiotic bouquet formation and efficient homologous synapsis in maize (*Zea mays* L.). Genetics **162**, 1979–1993.

**González RM, Ricardi MM, Iusem ND.** 2011. Atypical epigenetic mark in an atypical location: cytosine methylation at asymmetric (CNN) sites within the body of a non-repetitive tomato gene. BMC Plant Biology **11**, 1–11.

**González RM, Ricardi MM, Iusem ND.** 2013. Epigenetic marks in an adaptive water stress-responsive gene in tomato roots under normal and drought conditions. Epigenetics **8**, 864–872.

**Gourcilleau D, Bogeat-Triboulot M-B, Le Thiec D, Lafon-Placette C, Delaunay A, El-Soud WA, Brignolas F, Maury S.** 2010. DNA methylation and histone acetylation: genotypic variations in hybrid poplars, impact of

water deficit and relationships with productivity. Annals of Forest Science **67**, 208.

**Grob S, Schmid MW, Grossniklaus U.** 2014. Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Molecular Cell **55**, 678–693.

**Grossniklaus U, Kelly WG, Kelly B, Ferguson-Smith AC, Pembrey M, Lindquist S.** 2013. Transgenerational epigenetic inheritance: how important is it? Nature Reviews Genetics **14**, 228–235.

**Guo S, Sun B, Looi LS, Xu Y, Gan ES, Huang J, Ito T.** 2015. Co-ordination of flower development through epigenetic regulation in two model species: rice and Arabidopsis. Plant & Cell Physiology **56**, 830–842.

**Haag JR, Brower-Toland B, Krieger EK, et al.** 2014. Functional diversification of maize RNA polymerase IV and V subtypes via alternative catalytic subunits. Cell Reports **9**, 378–390.

**He Y.** 2012. Chromatin regulation of flowering. Trends in Plant Science **17**, 556–562.

**He Y, Wang M, Dukowic-Schulze S, et al.** 2017. Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. Proceedings of the National Academy of Sciences, USA **114**, 12231–12236.

**Heitz E.** 1928. Das Heterochromatin der Moose. Jahrbücher für Wissenschaftliche Botanik **69**, 762–818.

**Higgins JD, Osman K, Jones GH, Franklin FC.** 2014. Factors underlying restricted crossover localization in barley meiosis. Annual Review of Genetics **48**, 29–47.

**Higgins JD, Perry RM, Barakate A, Ramsay L, Waugh R, Halpin C, Armstrong SJ, Franklin FC.** 2012. Spatiotemporal asymmetry of the meiotic program underlies the predominantly distal distribution of meiotic crossovers in barley. The Plant Cell **24**, 4096–4109.

**Hirano T, Kobayashi R, Hirano M.** 1997. Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. Cell **89**, 511–521.

**Houben A, Kumke K, Nagaki K, Hause G.** 2011. CENH3 distribution and differential chromatin modifications during pollen development in rye (*Secale cereale* L.). Chromosome Research **19**, 471–480.

**Houben A, Wako T, Furushima-Shimogawara R, Presting G, Künzel G, Schubert I, Fukui K.** 2002. The cell cycle dependent phosphorylation of histone H3 is correlated with the condensation of plant mitotic chromosomes. The Plant Journal **18**, 675–679.

**Hsieh PH, He S, Buttress T, Gao H, Couchman M, Fischer RL, Zilberman D, Feng X.** 2016. Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. Proceedings of the National Academy of Sciences, USA **113**, 15132–15137.

**Hu Y, Morota G, Rosa GJ, Gianola D.** 2015. Prediction of plant height in *Arabidopsis thaliana* using DNA methylation data. Genetics **201**, 779–793.

**Hyun Y, Yun H, Park K, Ohr H, Lee O, Kim DH, Sung S, Choi Y.** 2013. The catalytic subunit of Arabidopsis DNA polymerase α ensures stable maintenance of histone modification. Development **140**, 156–166.

**Ibarra CA, Feng X, Schoft VK, et al.** 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science **337**, 1360–1364.

**Ingouff M, Berger F.** 2010. Histone3 variants in plants. Chromosoma **119**, 27–33.

**Ingouff M, Hamamura Y, Gourgues M, Higashiyama T, Berger F.** 2007. Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. Current Biology **17**, 1032–1037.

**Ingouff M, Rademacher S, Holec S, Soljić L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F.** 2010. Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. Current Biology **20**, 2137–2143.

**Internatioanl Barley Genome Sequencing Consortium**. 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature **491**, 711–716.

**Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J.** 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature **472**, 115–119.

**Iwasaki M, Paszkowski J.** 2014. Epigenetic memory in plants. The EMBO Journal **33**, 1987–1998.

**Jablonka E, Raz G.** 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Quarterly Review of Biology **84**, 131–176.

**Jackel JN, Storer JM, Coursey T, Bisaro DM.** 2016. Arabidopsis RNA polymerases IV and V are required to establish H3K9 methylation, but not cytosine methylation, on geminivirus chromatin. Journal of Virology **90**, 7529–7540.

**Jégu T, Latrasse D, Delarue M, et al.** 2013. Multiple functions of Kip-related protein5 connect endoreduplication and cell elongation. Plant Physiology **161**, 1694–1705.

**Jiang D, Berger F.** 2017. DNA replication-coupled histone modification maintains Polycomb gene silencing in plants. Science **357**, 1146–1149.

**Johnson C, Conrad LJ, Patel R, Anderson S, Li C, Pereira A, Sundaresan V.** 2018. Reproductive long intergenic noncoding RNAs exhibit male gamete specificity and polycomb repressive complex 2-mediated repression. Plant Physiology **177**, 1198–1217.

**Karimi-Ashtiyani R, Ishii T, Niessen M, et al.** 2015. Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proceedings of the National Academy of Sciences, USA **112**, 11211–11216.

**Kaszás E, Cande WZ.** 2000. Phosphorylation of histone H3 is correlated with changes in the maintenance of sister chromatid cohesion during meiosis in maize, rather than the condensation of the chromatin. Journal of Cell Science **113**, 3217–3226.

**Kawakatsu T, Nery JR, Castanon R, Ecker JR.** 2017. Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biology **18**, 171.

**Kawashima T, Berger F.** 2014. Epigenetic reprogramming in plant sexual reproduction. Nature Reviews Genetics **15**, 613–624.

**Kim JM, To TK, Nishioka T, Seki M.** 2010. Chromatin regulation functions in plant abiotic stress responses. Plant, Cell & Environment **33**, 604–611.

**Kotak S, Larkindale J, Lee U, von Koskull-Döring P, Vierling E, Scharf KD.** 2007. Complexity of the heat stress response in plants. Current Opinion in Plant Biology **10**, 310–316.

**Koyama M, Kurumizaka H.** 2018. Structural diversity of the nucleosome. Journal of Biochemistry **163**, 85–95.

**Kumar S, Beena AS, Awana M, Singh A.** 2017. Salt-induced tissue-specific cytosine methylation downregulates expression of HKT genes in contrasting wheat (*Triticum aestivum* L.) genotypes. DNA and Cell Biology **36**, 283–294.

**Kumar SV, Wigge PA.** 2010. H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell **140**, 136–147.

**Kwon HJ, Kim JH, Kim M, Lee JK, Hwang WS, Kim DY.** 2003. Antiparasitic activity of depudecin on *Neospora caninum* via the inhibition of histone deacetylase. Veterinary Parasitology **112**, 269–276.

**Lämke J, Bäurle I.** 2017. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biology **18**, 124.

**Latrasse D, Jégu T, Li H, et al.** 2017*a*. MAPK-triggered chromatin reprogramming by histone deacetylase in plant innate immunity. Genome Biology **18**, 131.

**Latrasse D, Rodriguez-Granados NY, Veluchamy A, et al.** 2017*b*. The quest for epigenetic regulation underlying unisexual flower development in *Cucumis melo*. Epigenetics & Chromatin **10**, 22.

**Lauss K, Wardenaar R, Oka R, van Hulten MHA, Guryev V, Keurentjes JJB, Stam M, Johannes F.** 2018. Parental DNA methylation states are associated with heterosis in epigenetic hybrids. Plant Physiology **176**, 1627–1645.

**Law JA, Jacobsen SE.** 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics **11**, 204–220.

**Lermontova I, Rutten T, Schubert I.** 2011. Deposition, turnover, and release of CENH3 at Arabidopsis centromeres. Chromosoma **120**, 633–640.

**Li G, Wang D, Yang R, et al.** 2014. Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. Proceedings of the National Academy of Sciences, USA **111**, 7582–7587.

**Li WQ, Zhang MJ, Gan PF, et al.** 2017. CLD1/SRL1 modulates leaf rolling by affecting cell wall formation, epidermis integrity and water homeostasis in rice. The Plant Journal **92**, 904–923.

**Liu C, Cheng YJ, Wang JW, Weigel D.** 2017. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nature Plants **3**, 742–748.

**Liu C, Moschou PN.** 2018. Phenotypic novelty by CRISPR in plants. Developmental Biology **435**, 170–175.

**Liu C, Wu D, Wang L, Dang J, He Q, Guo Q, Liang G.** 2018. Cis-regulated additively expressed genes play a fundamental role in the formation of triploid loquat (*Eriobotrya japonica* (Thunb.) Lindl.) heterosis. Molecular Genetics and Genomics **293**, 967–981.

**Liu H, Nonomura KI.** 2016. A wide reprogramming of histone H3 modifications during male meiosis I in rice is dependent on the Argonaute protein MEL1. Journal of Cell Science **129**, 3553–3561.

**Liu J, Tang X, Gao L, et al**. 2012. A role of tomato UV-damaged DNA binding protein 1 (DDB1) in organ size control via an epigenetic manner. PLoS One **7**, e42621.

**Liu M, Shi S, Zhang S, Xu P, Lai J, Liu Y, Yuan D, Wang Y, Du J, Yang C.** 2014. SUMO E3 ligase AtMMS21 is required for normal meiosis and gametophyte development in Arabidopsis. BMC Plant Biology **14**, 153.

**Liu Z, Makaroff CA.** 2006. Arabidopsis separase AESP is essential for embryo development and the release of cohesin during meiosis. The Plant Cell **18**, 1213–1225.

**Long JC, Xia AA, Liu JH, Jing JL, Wang YZ, Qi CY, He Y.** 2019. Decrease in DNA methylation 1 (DDM1) is required for the formation of m CHH islands in maize. Journal of Integrative Plant Biology **61**, 749–764.

**Lorković ZJ, Park C, Goiser M, Jiang D, Kurzbauer MT, Schlögelhofer P, Berger F.** 2017. Compartmentalization of DNA damage response between heterochromatin and euchromatin is mediated by distinct H2A histone variants. Current Biology **27**, 1192–1199.

**Maher B.** 2008. Personal genomes: the case of the missing heritability. Nature **456**, 18–21.

**Maheshwari S, Tan EH, West A, Franklin FC, Comai L, Chan SW.** 2015. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genetics **11**, e1004970.

**Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB.** 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics **38**, 948–952.

**Marand AP, Jansky SH, Zhao H, et al**. 2017. Meiotic crossovers are associated with open chromatin and enriched with Stowaway transposons in potato. Genome Biology **18**, 203.

**Marfil CF, Asurmendi S, Masuelli RW.** 2012. Changes in micro RNA expression in a wild tuber-bearing Solanum species induced by 5-azacytidine treatment. Plant Cell Reports **31**, 1449–1461.

**Martínez G, Panda K, Köhler C, Slotkin RK.** 2016. Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. Nature Plants **2**, 16030.

**Martinez G, Wolff P, Wang Z, Moreno-Romero J, Santos-González J, Conze LL, DeFraia C, Slotkin RK, Köhler C.** 2018. Paternal easiRNAs regulate parental genome dosage in Arabidopsis. Nature Genetics **50**, 193–198.

**Martinez-Garcia M, Schubert V, Osman K, Darbyshire A, Sanchez-Moran E, Franklin FCH.** 2018. TOPII and chromosome movement help remove interlocks between entangled chromosomes during meiosis. Journal of Cell Biology **217**, 4070–4079.

**Mascher M, Gundlach H, Himmelbach A, et al**. 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature **544**, 427–433.

**Mathieu O, Jasencakova Z, Vaillant I, Gendrel AV, Colot V, Schubert I, Tourmente S.** 2003. Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in Arabidopsis. The Plant Cell **15**, 2929–2939.

**Matsumoto M, Matsutani S, Sugita K, Yoshida H, Hayashi F, Terui Y, Nakai H, Uotani N, Kawamura Y, Matsumoto K.** 1992. Depudecin: a novel compound inducing the flat phenotype of NIH3T3 cells doubly transformed by ras- and src-oncogene, produced by *Alternaria brassicicola*. Journal of Antibiotics **45**, 879–885.

**Matzke MA, Mosher RA.** 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Reviews. Genetics **15**, 394–408.

**McGinty RK, Tan S.** 2015. Nucleosome structure and function. Chemical Reviews **115**, 2255–2273.

**Mehta GD, Rizvi SM, Ghosh SK.** 2012. Cohesin: a guardian of genome integrity. Biochimica et Biophysica Acta **1823**, 1324–1342.

**Mérai Z, Chumak N, García-Aguilar M, et al**. 2014. The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. Proceedings of the National Academy of Sciences, USA **111**, 16166–16171.

**Mercier R, Vezon D, Bullier E, Motamayor JC, Sellier A, Lefèvre F, Pelletier G, Horlow C.** 2001. SWITCH1 (SWI): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. Genes & Development **15**, 1859–1871.

**Mittelsten Scheid O, Afsar K, Paszkowski J.** 2003. Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. Nature Genetics **34**, 450–454.

**Moore JW, Loake GJ, Spoel SH.** 2011. Transcription dynamics in plant immunity. The Plant Cell **23**, 2809–2820.

**Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC.** 2009. Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature **460**, 283–286.

**Nasmyth K.** 2001. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annual Review of Genetics **35**, 673–745.

**Nasuda S, Hudakova S, Schubert I, Houben A, Endo TR.** 2005. Stable barley chromosomes without centromeric repeats. Proceedings of the National Academy of Sciences, USA **102**, 9842–9847.

**Okada T, Endo M, Singh MB, Bhalla PL.** 2005. Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. The Plant Journal **44**, 557–568.

**Oliver C, Pradillo M, Corredor E, Cuñado N.** 2013. The dynamics of histone H3 modifications is species-specific in plant meiosis. Planta **238**, 23–33.

**Oliver SN, Finnegan EJ, Dennis ES, Peacock WJ, Trevaskis B.** 2009. Vernalization-induced flowering in cereals is associated with changes in histone methylation at the VERNALIZATION1 gene. Proceedings of the National Academy of Sciences, USA **106**, 8386–8391.

**Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP.** 2010. Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature **464**, 628–632.

**Ong-Abdullah M, Ordway JM, Jiang N, et al**. 2015. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. Nature **525**, 533–537.

**Ono T, Fang Y, Spector DL, Hirano T.** 2004. Spatial and temporal regulation of Condensins I and II in mitotic chromosome assembly in human cells. Molecular Biology of the Cell **15**, 3296–3308.

**Palmgren MG, Edenbrandt AK, Vedel SE, et al**. 2015. Are we ready for back-to-nature crop breeding? Trends in Plant Science **20**, 155–164.

**Pandey P, Houben A, Kumlehn J, Melzer M, Rutten T.** 2013. Chromatin alterations during pollen development in *Hordeum vulgare*. Cytogenetic and Genome Research **141**, 50–57.

**Pecinka A, Abdelsamad A, Vu GT.** 2013. Hidden genetic nature of epigenetic natural variation in plants. Trends in Plant Science **18**, 625–632.

**Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O.** 2010. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. The Plant Cell **22**, 3118–3129.

**Pecinka A, Liu CH.** 2014. Drugs for plant chromosome and chromatin research. Cytogenetic and Genome Research **143**, 51–59.

**Pecinka A, Mittelsten Scheid O.** 2012. Stress-induced chromatin changes: a critical view on their heritability. Plant & Cell Physiology **53**, 801–808.

**Pecinka A, Schubert V, Meister A, Kreth G, Klatte M, Lysak MA, Fuchs J, Schubert I.** 2004. Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. Chromosoma **113**, 258–269.

**Périlleux C, Pieltain A, Jacquemin G, et al**. 2013. A root chicory MADS box sequence and the Arabidopsis flowering repressor FLC share common features that suggest conserved function in vernalization and de-vernalization responses. The Plant Journal **75**, 390–402.

**Phillips D, Nibau C, Wnetrzak J, Jenkins G.** 2012. High resolution analysis of meiotic chromosome structure and behaviour in barley (*Hordeum vulgare* L.). PLoS One **7**, e39539.

**Pietzenuk B, Markus C, Gaubert H, Bagwan N, Merotto A, Bucher E, Pecinka A.** 2016. Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. Genome Biology **17**, 209.

**Pignatta D, Novitzky K, Satyaki PRV, Gehring M.** 2018. A variably imprinted epiallele impacts seed development. PLoS Genetics **14**, e1007469.

**Pillot M, Baroux C, Vazquez MA, Autran D, Leblanc O, Vielle-Calzada JP, Grossniklaus U, Grimanelli D.** 2010. Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis. The Plant Cell **22**, 307–320.

**Porto DD, Bruneau M, Perini P, Anzanello R, Renou JP, dos Santos HP, Fialho FB, Revers LF.** 2015. Transcription profiling of the chilling requirement for bud break in apples: a putative role for FLC-like genes. Journal of Experimental Botany **66**, 2659–2672.

**Pradillo M, Knoll A, Oliver C, Varas J, Corredor E, Puchta H, Santos JL.** 2015. Involvement of the Cohesin cofactor PDS5 (SPO76) during meiosis and DNA repair in *Arabidopsis thaliana*. Frontiers in Plant Science **6**, 1034.

**Prusicki MA, Keizer EM, van Rosmalen RP, Komaki S, Seifert F, Müller K, Wijnker E, Fleck C, Schnittger A.** 2019. Live cell imaging of meiosis in *Arabidopsis thaliana*. eLife **8**, e42834.

**Quadrana L, Colot V.** 2016. Plant transgenerational epigenetics. Annual Review of Genetics **50**, 467–491.

**Raja P, Sanville BC, Buchmann RC, Bisaro DM.** 2008. Viral genome methylation as an epigenetic defense against geminiviruses. Journal of Virology **82**, 8997–9007.

**Rakyan VK, Down TA, Balding DJ, Beck S.** 2011. Epigenome-wide association studies for common human diseases. Nature Reviews. Genetics **12**, 529–541.

**Ramírez-González RH, Borrill P, Lang D, *et al*.** 2018. The transcriptional landscape of polyploid wheat. Science **361**, eaar6089.

**Ramirez-Prado JS, Piquerez SJM, Bendahmane A, Hirt H, Raynaud C, Benhamed M.** 2018. Modify the histone to win the battle: chromatin dynamics in plant–pathogen interactions. Frontiers in Plant Science **9**, 355.

**Ravi M, Chan SW.** 2010. Haploid plants produced by centromere-mediated genome elimination. Nature **464**, 615–618.

**Reeves PH, Ellis CM, Ploense SE, *et al*.** 2012. A regulatory network for coordinated flower maturation. PLoS Genetics **8**, e1002506.

**Richards DM, Greer E, Martin AC, Moore G, Shaw PJ, Howard M.** 2012. Quantitative dynamics of telomere bouquet formation. PLoS Computational Biology **8**, e1002812.

**Roberts NY, Osman K, Armstrong SJ.** 2009. Telomere distribution and dynamics in somatic and meiotic nuclei of *Arabidopsis thaliana*. Cytogenetic and Genome Research **124**, 193–201.

**Rockmill B, Voelkel-Meiman K, Roeder GS.** 2006. Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in *Saccharomyces cerevisiae*. Genetics **174**, 1745–1754.

**Roessler K, Bousios A, Meca E, Gaut BS.** 2018. Modeling interactions between transposable elements and the plant epigenetic response: a surprising reliance on element retention. Genome Biology and Evolution **10**, 803–815.

**Rosa S, Shaw P.** 2013. Insights into chromatin structure and dynamics in plants. Biology **2**, 1378–1410.

**Roudier F, Ahmed I, Bérard C, *et al*.** 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. The EMBO Journal **30**, 1928–1938.

**Ruelens P, de Maagd RA, Proost S, Theißen G, Geuten K, Kaufmann K.** 2013. FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. Nature Communications **4**, 2280.

**Russell SD, Gou X, Wong CE, Wang X, Yuan T, Wei X, Bhalla PL, Singh MB.** 2012. Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. New Phytologist **195**, 560–573.

**Rutowicz K, Puzio M, Halibart-Puzio J, *et al*.** 2015. A specialized histone H1 variant is required for adaptive responses to complex abiotic stress and related DNA methylation in Arabidopsis. Plant Physiology **169**, 2080–2101.

**Sandhu D, Gill KS.** 2002. Gene-containing regions of wheat and the other grass genomes. Plant Physiology **128**, 803–811.

**Sanei M, Pickering R, Kumke K, Nasuda S, Houben A.** 2011. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proceedings of the National Academy of Sciences, USA **108**, E498–E505.

**Schmiesing JA, Gregson HC, Zhou S, Yokomori K.** 2000. A human condensin complex containing hCAP-C-hCAP-E and CNAP1, a homolog of Xenopus XCAP-D2, colocalizes with phosphorylated histone H3 during the early stage of mitotic chromosome condensation. Molecular and Cellular Biology **20**, 6996–7006.

**Schmitz RJ, He Y, Valdés-López O, *et al*.** 2013*a*. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Research **23**, 1663–1674.

**Schmitz RJ, Schultz MD, Urich MA, *et al*.** 2013*b*. Patterns of population epigenomic diversity. Nature **495**, 193–198.

**Schoft VK, Chumak N, Mosiolek M, Slusarz L, Komnenovic V, Brownfield L, Twell D, Kakutani T, Tamaru H.** 2009. Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Reports **10**, 1015–1021.

**Schubert V, Klatte M, Pecinka A, Meister A, Jasencakova Z, Schubert I.** 2006. Sister chromatids are often incompletely aligned in meristematic and endopolyploid interphase nuclei of *Arabidopsis thaliana*. Genetics **172**, 467–475.

**Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C.** 2014. The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. The Plant Cell **26**, 2351–2366.

**Shao T, Tang D, Wang K, Wang M, Che L, Qin B, Yu H, Li M, Gu M, Cheng Z.** 2011. OsREC8 is essential for chromatid cohesion and metaphase I monopolar orientation in rice meiosis. Plant Physiology **156**, 1386–1396.

**She W, Grimanelli D, Rutowicz K, Whitehead MW, Puzio M, Kotlinski M, Jerzmanowski A, Baroux C.** 2013. Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. Development **140**, 4008–4019.

**Sheehan MJ, Pawlowski WP.** 2009. Live imaging of rapid chromosome movements in meiotic prophase I in maize. Proceedings of the National Academy of Sciences, USA **106**, 20989–20994.

**Shen X, Xu L, Liu Y, Dong H, Zhou D, Zhang Y, Lin S, Cao J, Huang L.** 2019. Comparative transcriptome analysis and ChIP-sequencing reveals stage-specific gene expression and regulation profiles associated with pollen wall formation in *Brassica rapa*. BMC Genomics **20**, 264.

**Shen Y, Zhang J, Liu Y, Liu S, Liu Z, Duan Z, Wang Z, Zhu B, Guo YL, Tian Z.** 2018. DNA methylation footprints during soybean domestication and improvement. Genome Biology **19**, 128.

**Shi J, Dong A, Shen WH.** 2014. Epigenetic regulation of rice flowering and reproduction. Frontiers in Plant Science **5**, 803.

**Shroff R, Arbel-Eden A, Pilch D, Ira G, Bonner WM, Petrini JH, Haber JE, Lichten M.** 2004. Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. Current Biology **14**, 1703–1711.

**Sidhu GK, Fang C, Olson MA, Falque M, Martin OC, Pawlowski WP.** 2015. Recombination patterns in maize reveal limits to crossover homeostasis. Proceedings of the National Academy of Sciences, USA **112**, 15982–15987.

**Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, Martienssen RA.** 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136**, 461–472.

**Smith SJ, Osman K, Franklin FC.** 2014. The condensin complexes play distinct roles to ensure normal chromosome morphogenesis during meiotic division in Arabidopsis. The Plant Journal **80**, 255–268.

**Song Q, Zhang T, Stelly DM, Chen ZJ.** 2017. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biology **18**, 99.

**Song T, Zhang Q, Wang H, Han J, Xu Z, Yan S, Zhu Z.** 2018. OsJMJ703, a rice histone demethylase gene, plays key roles in plant development and responds to drought stress. Plant Physiology and Biochemistry **132**, 183–188.

**Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H.** 2002. Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. Journal of Biological Chemistry **277**, 37741–37746.

**Stroud H, Otero S, Desvoyes B, Ramírez-Parra E, Jacobsen SE, Gutierrez C.** 2012. Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences, USA **109**, 5370–5375.

**Takuno S, Ran JH, Gaut BS.** 2016. Evolutionary patterns of genic DNA methylation vary across land plants. Nature Plants **2**, 15222.

**Teyssier E, Bernacchia G, Maury S, How Kit A, Stammitti-Bert L, Rolin D, Gallusci P.** 2008. Tissue dependent variations of DNA methylation and endoreduplication levels during tomato fruit development and ripening. Planta **228**, 391–399.

**Thieme M, Lanciano S, Balzergue S, Daccord N, Mirouze M, Bucher E.** 2017. Inhibition of RNA polymerase II allows controlled mobilisation of retrotransposons for plant breeding. Genome Biology **18**, 134.

**Tian L, Fong MP, Wang JJ, Wei NE, Jiang H, Doerge RW, Chen ZJ.** 2005. Reversible histone acetylation and deacetylation mediate genome-wide, promoter-dependent and locus-specific changes in gene expression during plant development. Genetics **169**, 337–345.

**Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I.** 2010. Stress-induced activation of heterochromatic transcription. PLoS Genetics **6**, e1001175.

**Trujillo JT, Seetharam AS, Hufford MB, Beilstein MA, Mosher RA.** 2018. Evidence for a unique DNA-dependent RNA polymerase in cereal crops. Molecular Biology and Evolution **35**, 2454–2462.

**Underwood CJ, Choi K, Lambing C, et al.** 2018. Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. Genome Research **28**, 519–531.

**van Esse GW, Walla A, Finke A, Koornneef M, Pecinka A, von Korff M.** 2017. Six-rowed Spike3 (VRS3) is a histone demethylase that controls lateral spikelet development in barley. Plant Physiology **174**, 2397–2408.

**Vanyushin BF, Alexandrushkina NI, Kirnos MD.** 1988. N6-Methyladenine in mitochondrial DNA of higher plants. FEBS Letters **233**, 397–399.

**Venetsky A, Levy-Zamir A, Khasdan V, Domb K, Kashkush K.** 2015. Structure and extent of DNA methylation-based epigenetic variation in wild emmer wheat (*T. turgidum* ssp. *dicoccoides*) populations. BMC Plant Biology **15**, 200.

**Verkest A, Byzova M, Martens C, et al.** 2015. Selection for improved energy use efficiency and drought tolerance in canola results in distinct transcriptome and epigenome changes. Plant Physiology **168**, 1338–1350.

**Vlot AC, Dempsey DA, Klessig DF.** 2009. Salicylic acid, a multifaceted hormone to combat disease. Annual Review of Phytopathology **47**, 177–206.

**Wako T, Houben A, Furushima-Shimogawara R, Belyaev ND, Fukui K.** 2003. Centromere-specific acetylation of histone H4 in barley detected through three-dimensional microscopy. Plant Molecular Biology **51**, 533–541.

**Wako T, Murakami Y, Fukui K.** 2005. Comprehensive analysis of dynamics of histone H4 acetylation in mitotic barley cells. Genes & Genetic Systems **80**, 269–276.

**Walker J, Gao H, Zhang J, Aldridge B, Vickers M, Higgins JD, Feng X.** 2018. Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. Nature Genetics **50**, 130–137.

**Wang C, Gao F, Wu J, Dai J, Wei C, Li Y.** 2010. Arabidopsis putative deacetylase AtSRT2 regulates basal defense by suppressing PAD4, EDS5 and SID2 expression. Plant & Cell Physiology **51**, 1291–1299.

**Wang LC, Wu JR, Hsu YJ, Wu SJ.** 2015. Arabidopsis HIT4, a regulator involved in heat-triggered reorganization of chromatin and release of transcriptional gene silencing, relocates from chromocenters to the nucleolus in response to heat stress. New Phytologist **205**, 544–554.

**Wang M, Tang D, Wang K, Shen Y, Qin B, Miao C, Li M, Cheng Z.** 2011. OsSGO1 maintains synaptonemal complex stabilization in addition to protecting centromeric cohesion during rice meiosis. The Plant Journal **67**, 583–594.

**Wang Z, Casas-Mollano JA, Xu J, Riethoven JJ, Zhang C, Cerutti H.** 2015. Osmotic stress induces phosphorylation of histone H3 at threonine 3 in pericentromeric regions of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences, USA **112**, 8487–8492.

**Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer NM.** 2013. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences, USA **110**, 19639–19644.

**Whittaker C, Dean C.** 2017. The FLC locus: a platform for discoveries in epigenetics and adaptation. Annual Review of Cell and Developmental Biology **33**, 555–575.

**Willing EM, Piofczyk T, Albert A, Winkler JB, Schneeberger K, Pecinka A.** 2016. UVR2 ensures transgenerational genome stability under simulated natural UV-B in *Arabidopsis thaliana*. Nature Communications **7**, 13522.

**Wollmann H, Holec S, Alden K, Clarke ND, Jacques PÉ, Berger F.** 2012. Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. PLoS Genetics **8**, e1002658.

**Xianwei S, Zhang X, Sun J, Cao X.** 2015. Epigenetic mutation of RAV6 affects leaf angle and seed size in rice. Plant Physiology **169**, 2118–2128.

**Xie N, Zhou Y, Sun Q, Tang B.** 2018. Novel epigenetic techniques provided by the CRISPR/Cas9 system. Stem Cells International **2018**, 7834175.

**Xu J, Tanino KK, Robinson SJ.** 2016. Stable epigenetic variants selected from an induced hypomethylated *Fragaria vesca* population. Frontiers in Plant Science **7**, 1768.

**Yang C, Hamamura Y, Sofroni K, Böwer F, Stolze SC, Nakagami H, Schnittger A.** 2019. SWITCH 1/DYAD is a WINGS APART-LIKE antagonist that maintains sister chromatid cohesion in meiosis. Nature Communications **10**, 1755.

**Yang H, Liu X, Xin M, Du J, Hu Z, Peng H, Rossi V, Sun Q, Ni Z, Yao Y.** 2016. Genome-wide mapping of targets of maize histone deacetylase HDA101 reveals its function and regulatory mechanism during seed development. The Plant Cell **28**, 629–645.

**Yang X, Boateng KA, Strittmatter L, Burgess R, Makaroff CA.** 2009. Arabidopsis separase functions beyond the removal of sister chromatid cohesion during meiosis. Plant Physiology **151**, 323–333.

**Yang X, Timofejeva L, Ma H, Makaroff CA.** 2006. The Arabidopsis SKP1 homolog ASK1 controls meiotic chromosome remodeling and release of chromatin from the nuclear membrane and nucleolus. Journal of Cell Science **119**, 3754–3763.

**Yelagandula R, Stroud H, Holec S, et al.** 2014. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell **158**, 98–109.

**Yuan G, Ahootapeh BH, Komaki S, Schnittger A, Lillo C, De Storme N, Geelen D.** 2018. PROTEIN PHOSHATASE 2A B'α and β maintain centromeric sister chromatid cohesion during meiosis in Arabidopsis. Plant Physiology **178**, 317–328.

**Yuan J, Guo X, Hu J, Lv Z, Han F.** 2015. Characterization of two CENH3 genes and their roles in wheat evolution. New Phytologist **206**, 839–851.

**Yuan L, Yang X, Makaroff CA.** 2011. Plant cohesins, common themes and unique roles. Current Protein & Peptide Science **12**, 93–104.

**Zamariola L, De Storme N, Vannerum K, Vandepoele K, Armstrong SJ, Franklin FC, Geelen D.** 2014. SHUGOSHINs and PATRONUS protect meiotic centromere cohesion in *Arabidopsis thaliana*. The Plant Journal **77**, 782–794.

**Zelkowski M, Zelkowska K, Conrad U, Hesse S, Lermontova I, Marzec M, Meister A, Houben A, Schubert V.** 2019. Arabidopsis NSE4 proteins act in somatic nuclei and meiosis to ensure plant viability and fertility. Frontiers in Plant Science **10**, 774.

**Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D.** 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell **153**, 193–205.

**Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D.** 2010. Local DNA hypomethylation activates genes in rice endosperm. Proceedings of the National Academy of Sciences, USA **107**, 18729–18734.

**Zhang L, Cheng Z, Qin R, et al.** 2012. Identification and characterization of an epi-allele of FIE1 reveals a regulatory linkage between two epigenetic marks in rice. The Plant Cell **24**, 4407–4421.

**Zhang Q, Wang D, Lang Z, et al.** 2016. Methylation interactions in Arabidopsis hybrids require RNA-directed DNA methylation and are

influenced by genetic variation. Proceedings of the National Academy of Sciences, USA **113**, E4248–E4256.

**Zhang S, Wang D, Zhang H, Skaggs MI, Lloyd A, Ran D, An L, Schumaker KS, Drews GN, Yadegari R.** 2018. FERTILIZATION-INDEPENDENT SEED-polycomb repressive complex 2 plays a dual role in regulating type I MADS-box genes in early endosperm development. Plant Physiology **177**, 285–299.

**Zhang X.** 2012. Chromatin modifications in plants. In: Wendel FJ, Greilhuber J, Dolezel J, Leitch JI, eds. Plant genome diversity Vol 1: plant genomes, their residents, and their evolutionary dynamics. Vienna: Springer Vienna, 237–255.

**Zhang X, Li X, Marshall JB, Zhong CX, Dawe RK.** 2005. Phosphoserines on maize CENTROMERIC HISTONE H3 and histone H3 demarcate the centromere and pericentromere during chromosome segregation. The Plant Cell **17**, 572–583.

**Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, Qu LH, Shu WS, Chen YQ.** 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biology **15**, 512.

**Zhang YY, Latzel V, Fischer M, Bossdorf O.** 2018. Understanding the evolutionary potential of epigenetic variation: a comparison of heritable phenotypic variation in epiRILs, RILs, and natural ecotypes of *Arabidopsis thaliana*. Heredity **121**, 257–265.

**Zhong S, Fei Z, Chen YR, *et al*.** 2013. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nature Biotechnology **31**, 154–159.

**Zhou A, Pawlowski WP.** 2014. Regulation of meiotic gene expression in plants. Frontiers in Plant Science **5**, 413.

**Zhou C, Zhang L, Duan J, Miki B, Wu K.** 2005. HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in Arabidopsis. The Plant Cell **17**, 1196–1204.

**Zhu J, Fang L, Yu J, Zhao Y, Chen F, Xia G.** 2018. 5-Azacytidine treatment and TaPBF-D over-expression increases glutenin accumulation within the wheat grain by hypomethylating the Glu-1 promoters. Theoretical and Applied Genetics **131**, 735–746.

**Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S.** 2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456**, 125–129.

**Publikace 2**

# Comparative analysis of epigenetic inhibitors reveals different degrees of interference with transcriptional gene silencing and induction of DNA damage

Anna Nowicka[1,2,3], Barbara Tokarz[1,4], Jana Zwyrtková[1], Eva Dvořák Tomaštíková[1], Klára Procházková[1], Ugur Ercan[2], Andreas Finke[2], Wilfried Rozhon[5], Brigitte Poppenberger[5], Miroslav Otmar[6], Igor Niezgodzki[7], Marcela Krečmerová[6], Ingo Schubert[8] and Ales Pecinka[1,2,*] (iD)

[1]*Institute of Experimental Botany (IEB), Czech Academy of Sciences, Centre of the Region Haná for Biotechnological and Agricultural Research (CRH), CZ-779 00 Olomouc, Czech Republic,*
[2]*Max Planck Institute for Plant Breeding Research (MPIPZ), DE-50829 Cologne, Germany,*
[3]*The Polish Academy of Sciences, The Franciszek Górski Institute of Plant Physiology, Niezapominajek 21, PL-30 239 Krakow, Poland,*
[4]*Unit of Botany and Plant Physiology, Institute of Plant Biology and Biotechnology, Faculty of Biotechnology and Horticulture, University of Agriculture in Krakow, Al. 29 Listopada 54, PL-31 425 Krakow, Poland,*
[5]*Biotechnology of Horticultural Crops, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Liesel-Beckmann-Straße 1, DE-85354 Freising, Germany,*
[6]*Institute of Organic Chemistry and Biochemistry, CZ-166 10 Praha 6, Czech Republic,*
[7]*Biogeosystem Modelling Group, ING PAN – Institute of Geological Sciences Polish Academy of Sciences, Research Center in Krakow, Senacka 1, PL-31 002 Krakow, Poland, and*
[8]*Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, DE-06466 Gatersleben, OT, Germany*

## SUMMARY

**Repetitive DNA sequences and some genes are epigenetically repressed by transcriptional gene silencing (TGS). When genetic mutants are not available or problematic to use, TGS can be suppressed by chemical inhibitors. However, informed use of epigenetic inhibitors is partially hampered by the absence of any systematic comparison. In addition, there is emerging evidence that epigenetic inhibitors cause genomic instability, but the nature of this damage and its repair remain unclear. To bridge these gaps, we compared the effects of 5-azacytidine (AC), 2′-deoxy-5-azacytidine (DAC), zebularine and 3-deazaneplanocin A (DZNep) on TGS and DNA damage repair. The most effective inhibitor of TGS was DAC, followed by DZNep, zebularine and AC. We confirmed that all inhibitors induce DNA damage and suggest that this damage is repaired by multiple pathways with a critical role of homologous recombination and of the SMC5/6 complex. A strong positive link between the degree of cytidine analog-induced DNA demethylation and the amount of DNA damage suggests that DNA damage is an integral part of cytidine analog-induced DNA demethylation. This helps us to understand the function of DNA methylation in plants and opens the possibility of using epigenetic inhibitors in biotechnology.**

**Keywords: DNA methylation, genome stability, DNA damage, cytidine analog, epigenetic inhibitors, *Arabidopsis thaliana*, *Vicia faba*.**

## INTRODUCTION

Eukaryotic nuclear genomes are composed of linear DNA molecules (chromosomes), which are wrapped around histone octamers to form nucleosomes, i.e. the basic units of chromatin (Alberts, 2002). Nucleosome arrays are folded into chromatin fibers and domains, chromosome territories or individually distinguishable chromosomes during cell division (Liu and Weigel, 2015; Meier *et al.*, 2017). The biochemical properties and functions of chromatin are defined by an intricate network of epigenetic information stored at all levels of genomic organization.

The presence of a methyl group at the fifth position of the cytosine aromatic ring (hereafter called DNA methylation) is a prominent chromatin modification with diverse

functions in plants (Feng *et al.*, 2010; Law and Jacobsen, 2010). Three basic functional DNA methylation contexts distinguished in plants are CG, CHG and CHH (where H is A, T or G). DNA methylation exclusively in the CG context occurs in about one-third of protein-coding genes in the model plant *Arabidopsis thaliana* and seems to be associated with high transcriptional activity (Zilberman *et al.*, 2007; Kawakatsu *et al.*, 2016). However, the exact function of gene body methylation is still unclear, and some plant species lack it completely (Takuno and Gaut, 2012; Bewick *et al.*, 2016; Kawakatsu *et al.*, 2016). By contrast, accumulation of CG, CHG and CHH methylation in gene promoters and repetitive sequences suppresses transcription and leads to heterochromatinization, i.e. enrichment of chromatin with histone-repressive modifications and strong chromatin compaction, that is often observable as intensely stained nuclear/chromosomal regions (Fransz *et al.*, 2003; Lister *et al.*, 2008; Stroud *et al.*, 2013). DNA methylation in all sequence contexts can be established *de novo* in plants by the DOMAINS REARRANGED METHYLTRANSFERASE (DRM) family of DNA methyltransferases (DNMTs), which are directed to their sites of action by small, typically 24-nucleotide-long, double-stranded RNAs in the process of RNA-directed DNA methylation (RdDM) (Matzke and Mosher, 2014). In addition, a vast amount of CHH methylation at the termini of long transposons is established in Arabidopsis in a small-RNA-independent manner by CHROMOMETHYLASE 2 (CMT2) (Zemach *et al.*, 2013). Once established, DNA methylation can be perpetuated during DNA replication by the CG and CHG DNA methyltransferases DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), respectively (Mathieu *et al.*, 2007; Du *et al.*, 2012). Any potential flaws in the methylation patterns can be subsequently corrected by canonical RdDM (Baubec *et al.*, 2014).

Interference with DNA methylation is achieved by suppressing or mutating genes that control DNA methylation (Ossowski *et al.*, 2008; Fauser *et al.*, 2014; O'Malley *et al.*, 2015). In situations when genetic mutants are not available, suppression of the target gene(s) is not possible or only transient effects are needed, chemical inhibitors that interfere with DNA methylation and/or histone modifications (epigenetic inhibitors) offer a useful alternative (Lyko and Brown, 2005; Yoo and Jones, 2006; Pecinka and Liu, 2014). The most commonly used inhibitors in the plant field are the non-methylable cytidine analogs 5-azacytidine (AC) and zebularine (Zeb), the methyl group synthesis inhibitor 3-deazaneplanocin A (DZNep) and the histone deacetylase class I and II inhibitor trichostatin A (TSA). Epigenetic inhibitors were instrumental in understanding the dynamics of DNA methylation and transposon silencing in *A. thaliana*, *Nicotiana tabacum* and cereals (Fajkus *et al.*, 1992; Fojtová *et al.*, 1998; Kovarik *et al.*, 2000; Griffin *et al.*, 2016), understanding the mechanisms of establishment and

maintenance of DNA methylation (Baubec *et al.*, 2010, 2014) and altering the plant developmental program (Fulneček *et al.*, 2011; Solís *et al.*, 2015; Nowicka *et al.*, 2019). In addition, there are reports from prokaryotes, fungi, animals and plants that specific epigenetic inhibitors cause genomic instability (Zadražil *et al.*, 1965; Fučík *et al.*, 1970; Hegde *et al.*, 1996; Kiziltepe *et al.*, 2007; Kuo *et al.*, 2007; Cho *et al.*, 2011; Orta *et al.*, 2014; Liu *et al.*, 2015). By the 1970s, induction of chromatid breaks by AC and 2′-deoxy-5-azacytidine (DAC) had been observed in *Vicia faba* (Fučík *et al.*, 1970). Furthermore, treatment with Zeb led to rearrangements of mitotic chromosomes in wheat (Cho *et al.*, 2011). However, the nature of the inhibitor-induced DNA damage and its repair mechanism(s) remain unknown.

The DNA damage repair (DDR) pathways follow a common strategy. The occurrence of DNA damage is recognized by a sensor, which transmits the information through a signaling cascade to the effectors responsible for the repair and regulation of connected cellular processes (e.g. the cell cycle) (for reviews see, e.g., Kimura and Sakaguchi, 2006; Hu *et al.*, 2016). The major plant DDR pathways are represented by base excision repair (BER), nucleotide excision repair (NER), mismatch repair, photoreactivation, non-homologous end joining (NHEJ) and homologous recombination (HR). Recently, our laboratory revealed that ATAXIA TELANGIECTASIA MUTATED (ATM) and ATM AND RAD3-RELATED (ATR) DNA damage signaling kinases, but not DNA LIGASE 4 (LIG4) or KU70, are required for normal resistance to Zeb (Liu *et al.*, 2015), suggesting an important role of HR in the repair of Zeb-induced DNA damage. However, the contribution of other DDR pathways and the DNA-damaging effects of other inhibitors were not tested systematically.

The aim of our study is to perform a comprehensive comparison of epigenetic inhibitors with respect to their effects on plant nuclear morphology, DNA methylation and silencing of repetitive DNA. Furthermore, the induction of DNA damage by epigenetic inhibitors is still a little recognized effect and can sometimes even be confused with DNA demethylation effects. Therefore, we compared the DNA-damaging effects of specific epigenetic inhibitors and defined several repair pathways that are involved in mitigation of their genotoxic effects. Collectively, this will provide a better understanding of their mode of action and a more informed selection and evaluation of the phenotypes in future studies.

## RESULTS

### Comparison of epigenetic inhibitors in transcriptional gene silencing of a reporter gene

For the primary comparison, we selected nine known and/or potential epigenetic inhibitors representing three functionally diverse groups (Figure 1a): (i) the non-methylable

cytidine analogs AC, 5,6-dihydro-5-azacytidine (DHAC), DAC, α-2′-deoxy-5-azacytidine (α-DAC), 2′-deoxy-5,6-dihydro-5-azacytidine (DHDAC), α-2′-deoxy-5,6-dihydro-5-azacytidine (α-DHDAC) and Zeb; (ii) the S-ADENOSYL-L-HOMOCYSTEIN HYDROLASE (SAHH) inhibitor DZNep, which suppresses biosynthesis of the methyl group (Glazer *et al.*, 1986), and inhibits E(z)2, the catalytic subunit of the Polycomb Repressive Complex 2, in mammals (Fiskus *et al.*, 2009); (iii) TSA, the inhibitor of the class I and II histone deacetylases. In the direct treatment (DT), the plants were grown directly on the drug-containing media for 7 days and then analyzed (Figure 1b). In the postponed treatment (PT), they were first grown for 7 days on drug-free media and then on drug-containing media for another 7 days. The drugs were applied in concentrations of 5, 20 and 50 μM. As a readout for drug toxicity, we scored for the primary root length (Figure 2a,b and Figure S1 in the online Supporting Information). Minimal growth reduction was observed after the treatments with DHAC, DHDAC and α-DHDAC, intermediate reduction with AC and Zeb and strong reduction with DAC, α-DAC (only in the DT protocol), DZNep and TSA. It should be noted that DZNep strongly suppressed shoot growth over the root in DT when compared with other drugs (Figure 2a).

Next, we selected the drugs with the highest potential for reviving transcriptionally silenced genes by screening for activation of the transcriptionally silenced multi-copy GUS locus (*TsGUS*) (Morel *et al.*, 2000). *TsGUS* is fully silenced in L5 (6b5) wild-type (WT) plants, but strongly reactivated upon introduction into the mutant background of *DECREASED DNA METHYLATION1* (*DDM1*) chromatin remodeling factor (Elmayan *et al.*, 2005; Baubec *et al.*, 2014; Figures 2c,d and S2). After DT and PT, we observed intense GUS signals, comparable to those in *ddm1-5* plants, for all concentrations of DAC (Figures 2c,d and S2). This was followed by weaker staining (in decreasing order) in plants exposed to α-DAC (all concentrations), AC (50 μM), DZNep (5 and 20 μM) and Zeb (50 μM). Although DZNep led to strong staining in the shoot there was only minimal activation in the roots, and the total amount of GUS was lower than for DAC or L5 *ddm1* samples (Figures 2c,d and S2), probably due to a strong cytotoxic effect of this drug. No GUS staining was observed after the application of DHAC, DHDAC, α-DHDAC or, surprisingly, TSA (Figures 2c,d and S2). The amounts of GUS enzyme produced after some treatments were so high that it even cleaved the substrate in the surrounding staining solution and led to its coloration (Figure 2e), which provided another semi-quantitative readout of our experiment and pointed to the best candidates. This also showed that 50 μM DZNep was highly toxic for plants, as suggested by the reduced GUS signal (Figure 2e).

Based on these experiments, we considered AC, DAC (α-DAC), Zeb and lower concentrations of DZNep as the most promising drugs for interference with transcriptional gene silencing (TGS) and used them as the core set for further experiments.

## Tandem repetitive sequences show partial loss of DNA methylation upon inhibitor application

Epigenetic inhibitors reduce DNA methylation, but the existing data are not directly comparable due to many experimental variations between different studies. Therefore, we compared the DNA demethylation potential of AC, DAC, Zeb and DZNep in multiple assays. First, we immunostained 5-methyl-2′-deoxycytidine (5mdC) in isolated Arabidopsis nuclei using a specific antibody (Figure 3a). Wild-type mock-treated nuclei showed typical patterns with signals concentrated to heterochromatic chromocenters (CCs), i.e. intensely stained chromatin regions (Fransz *et al.*, 2003). After treatment with drug concentrations of 20 μM, the signals appeared more dispersed than in the mock-treated samples but not necessarily weaker. In order to obtain quantitative data, we determined the proportion of 5mdCs relative to all dCs in genomic DNA using high-performance liquid chromatography (Figure 3b). In DNA of mock-treated WT plants 6.1% of all dC residues were methylated, while only 3.2% (reduction to 52%) were methylated in *ddm1* plants. Treatment of WT plants with AC, Zeb and DZNep led to a reduction of approximately 20% in 5dmC compared with mock, and the reduction was even stronger (24.7%) after DAC treatment. To gain information on DNA methylation in a locus- and sequence-specific context, we performed Southern blots using genomic DNA of DT and PT plants digested with *Hpa*II, *Msp*I and *Alu*I (indicative of CG, CHG and CHH methylation, respectively) and probed with the Arabidopsis centromeric satellite (*pAL*) and *5S rDNA* repeats (Figure 3c and S3). Mock-treated WT and *ddm1* were used as high- and low-methylated controls, respectively. The most prominent demethylation was observed for a CG context, but none of the drugs reached the demethylation level of *ddm1* plants. Direct treatment was more effective than PT and the pattern was slightly different for each of the repeats. While *pAL* was strongly demethylated by DAC and DZNep, *5S rDNA* was more demethylated by AC and Zeb. Such differences could be related to the transcriptional activity of both loci and the possibility of incorporating non-metabolized AC and Zeb into RNA. For a CHG context we observed only minimal changes, except for direct treatment with DZNep, which reduced CHG methylation in both *pAL* and *5S rDNA*.

To assess transcriptional activation of specific targets controlled by TGS, we performed reverse transcription followed by quantitative PCR (RT-qPCR) on samples treated with 0 (mock) and 40 μM drugs for 48 h (Figure 3d). The maintenance methylation-silenced targets *TsGUS*, *TRANSCRIPTIONALLY SILENT INFORMATION* (*TSI*) and *LINE1-4*

(Steimer *et al.*, 2000; Baubec *et al.*, 2014), were strongly activated in the positive control L5 *ddm1-5*, moderately in DAC and weakly in AC-, Zeb- and DZNep-treated plants. The pattern of *TsGUS* mRNA levels corresponded well to the results of GUS staining. However, only some drug-induced changes were significant, possibly due to inter-experimental variation in drug treatment experiments. For the RdDM targets *soloLTR* and *SUPPRESSOR OF DRM1 DRM2 CMT3* (*SDC*) (Huettel *et al.*, 2006; Moissiard *et al.*, 2012), we found moderate activation in *L5 ddm1-5* and the same or even higher activation after DAC treatment. *SDC*, but not *soloLTR*, was partially activated also after AC, Zeb and DZNep treatments. The *Ta3* transposon, whose silencing is controlled mainly by H3K9me2 modification (Jackson *et al.*, 2002), was moderately transcriptionally activated in *L5 ddm1-5*, but only slightly after drug treatments, suggesting their primary effect is via DNA demethylation.

### Inhibitors reduce heterochromatin and lead to the dispersion of CCs

Some epigenetic inhibitors are known to affect chromatin organization, but information for others is missing. Therefore, we investigated the nuclear morphology after inhibitor treatment. First, we quantified the area of heterochromatin (represented by CCs) relative to the whole nuclear area. In mock-treated nuclei, CCs occupied 12.2% of the nuclear area, which is in agreement with published data (Soppe *et al.*, 2002). Treatment with AC, DAC, Zeb and DZNep significantly reduced the heterochromatin fraction to 9.6%, 8.6%, 9.6% and 8.4%, respectively, representing a 21.1–31.1% reduction relative to the control (Figure 4a,b). However, the CCs were still observable in most nuclei. Analysis of the organization of tandem repetitive DNA arrays represented by centromeric repeat (*pAL*) and *5S rDNA* by fluorescence *in situ* hybridization (FISH) revealed that their signals became partially or fully dispersed after epigenetic inhibitor treatment (Figure 4c–f). Collectively, reduced CC size and dispersion of repetitive DNA sequences suggest that the treatment with epigenetic inhibitors strongly affects heterochromatin organization. The most effective drugs tested here were DAC and DZNep.

### Epigenetic inhibitors induce chromosomal aberrations

Studies in various organisms indicated that treatment with epigenetic inhibitors leads to genomic instability. We analyzed the effects of the core set of epigenetic inhibitors on chromosome integrity and the cell cycle. Since Arabidopsis has small chromosomes, hampering clear resolution of structural changes, we used root apical meristems (RAMs) of 24-h mock- and drug-treated *V. faba* plants of genotype ACB with individually distinguishable chromosomes (Fučík *et al.*, 1970). First, we quantified nuclei representing $G_1$, S and $G_2$ phases of the cell cycle by flow cytometry

(Figure 5a,b). Mock treated RAMs contained 32.7%, 15.4% and 51.9% $G_1$, S and $G_2$ nuclei, respectively. The number of S-phase nuclei increased and that of $G_2$ decreased in response to AC, DAC and Zeb, possibly indicating problems with DNA replication (Figure 5c). By contrast, the vast majority of nuclei in DZNep-treated roots were in $G_2$ phase (66.5%), while S-phase nuclei were almost absent (Figure 5c). The frequency of mitoses decreased from 9.5% in the control plants to 6.5% and 6.0% in AC- and Zeb-treated RAMs and to only 3.2% and 2.9% in DAC- and DZNep-treated RAMs, respectively (Figure 5d). Next, we blocked the cells in metaphase using colchicine and analyzed chromosome aberrations (Figure 5e–h). Segment extensions (SEs), i.e. highly decondensed chromosomal regions (absent in mock-treated plants), were frequent in AC-treated (22%), moderately common in DAC- and Zeb-treated (7.5–9.3%), and rare in DZNep-treated (2.8%) plants. Breakage-based aberrations were dominated by isochromatid breaks (in 11.5% of metaphases of AC-treated, 17.1% of Zeb-treated and 22.6% of DAC-treated plants) (Figure 5f–h, Table S1). Structural chromatid aberrations occurred in 16.8% of metaphases of DZNep-treated plants and were represented similarly by reciprocal translocations, isochromatid breaks, interstitial deletions, single-chromatid breaks and unidentified structures (Figure 5g, Table S1). The highest frequency of isochromatid breaks was observed in the NOR region of chromosome III (Figure 5h, Table S2). At later stages of mitosis, we observed significantly increased (Tukey's test, $P \leq 0.05$) frequencies of anaphases with chromosomal bridges and micronuclei after the treatment with each of the inhibitors, indicating the presence of dicentric chromosomes and loss of genetic information (Figure 5i,j).

Hence, epigenetic inhibitors affect the cell cycle, reduce the number of cell divisions and induce segment extensions and breakage-based chromosome aberrations resulting in reduced genomic stability.

### Multiple pathways are involved in the repair of inhibitor-induced DNA damage

To shed light on the mechanism of inhibitor-induced DDR, we performed direct drug treatment followed by phenotypic analysis of 14 single or double mutants representing different DDR pathways: NER (*xpf*; Fidantsef *et al.*, 2000), BER (*ung*; Córdoba-Cañero *et al.*, 2010), NHEJ/HR (*atm*, *atr*, *atm atr*, *sog1*; Culligan *et al.*, 2006; Yoshiyama *et al.*, 2009), NHEJ (*ku70*, *lig4*; Riha *et al.*, 2002; van Attikum *et al.*, 2003), HR (*mus81*, *recq4a*; Schiml *et al.*, 2016), inter-strand crosslink repair (ICL; *fan1*; Herrmann *et al.*, 2015), DDR linked cell cycle control (*wee1*; De Schutter *et al.*, 2007) and SMC5/6-based repair (*smc6b*, *nse4a*; Mengiste *et al.*, 1999; Watanabe *et al.*, 2009; Diaz *et al.*, 2019). Treatment with 20 μM DZNep strongly suppressed shoot development of all tested genotypes but stimulated root growth in WT and *recq4a* plants (124% and 118% relative to mock treatment,
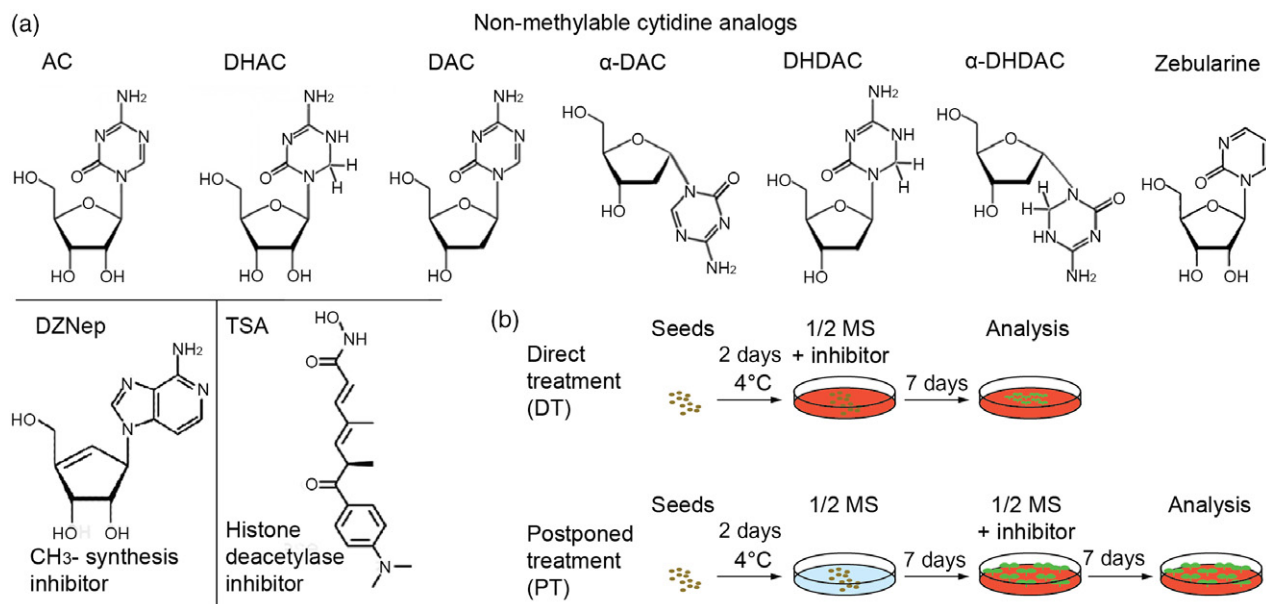
**Figure 1.** Epigenetic inhibitors and plant treatments used in this study.
(a) Chemical formulae and classification of the drugs used in this study. Non-methylable cytidine analogs: 5-azacytidine (AC), 5,6-dihydro-5-azacytidine (DHAC), 2′-deoxy-5-azacytidine (DAC), α-2′-deoxy-5-azacytidine (α-DAC), 2′-deoxy-5,6-dihydro-5-azacytidine (DHDAC), α-2′-deoxy-5,6-dihydro-5-azacytidine (α-DHDAC) and zebularine (Zeb). Methyl group (–CH₃) synthesis inhibitor 3-deazaneplanocin A (DZNep) and histone deacetylase inhibitor trichostatin A (TSA).
(b) Schematic representations of the direct treatment (DT) and the postponed treatment (PT) protocols. d = days, ½MS = half-strength Murashige–Skoog medium.

respectively; Figure 6a,b). The rest of the mutants had a root length reduced by about 10–30%, and the most sensitive genotypes (>40% growth reduction) were *atm atr* and *fan1*, indicating that the DZNep-induced damage is repaired predominantly by HR and ICL with a smaller contribution of other pathways. Zebularine treatment led to a 27% reduction in root length for WT, but up to a 65% reduction for *atm atr*, *mus81*, *smc6b* and *nse4a* plants. The phenotypes induced by AC were similar, but with a lower sensitivity of *mus81* plants. This trend was even more pronounced in the case of DAC treatment, where the single mutants *atm* and *atr* were already strongly hypersensitive (>80% reduced root length compared with 58% in the WT). In addition, we also observed strongly reduced growth of

*wee1*, *smc6b* and *nse4a* mutants on DAC-containing media. Partial sensitivity of *sog1* plants to DZNep and DAC and a WT-like response to Zeb and AC suggests that the inhibitor-induced damage is partially or fully SOG1-dependent, respectively. The role of the HR pathway was further reflected at the transcriptional level by increased transcription of the DDR markers *RADIATION SENSITIVE51* (*RAD51*), *RAD17* and *BREAST CANCER SUSCEPTIBILITY1* (*BRCA1*) in response to 24-h treatment with 20 μM 5-azacytidine analogs (Figure 6c).

The mutant analysis suggested that the HR pathway plays an important role in detoxification of the inhibitor-induced damage, and also some of the inhibitors were previously shown to enhance the frequency of HR in

**Figure 2.** Plant growth and *TsGUS* reporter locus activation in response to drug treatment.
(a) Representative phenotypes of wild type (WT) plants grown in the absence (mock) or the presence of 20 μM epigenetic inhibitors under direct treatment (DT, top) and postponed treatment (PT, bottom) regimes. Scale bars = 10 mm. Phenotypes of plants grown under the 5 and 50 μM concentrations are shown in Figure S1. AC, 5-azacytidine; DHAC, 5,6-dihydro-5-azacytidine; DAC, 2′-deoxy-5-azacytidine; α-DAC, α-2′-deoxy-5-azacytidine; DHDAC, 2′-deoxy-5,6-dihydro-5-azacytidine; α-DHDAC, α-2′-deoxy-5,6-dihydro-5-azacytidine; Zeb, zebularine; DZNep, 3-deazaneplanocin A; TSA, trichostatin A.
(b) Quantification of primary root length of plants grown under the DT (top) and the PT (bottom) regimes. Error bars represent standard deviation between the means of three biological replicates, each containing at least 25 plants. Letters in the bar area indicate statistically significantly different groups according to Tukey's test ($P \leq 0.05$).
(c) Representative activation patterns of the *TsGUS* locus in shoots in plants grown under DT conditions with 5, 20 and 50 μM concentrations of epigenetic inhibitors. L5 WT and L5 *ddm1* grown without inhibitors served as the silenced and the non-silenced *TsGUS* controls, respectively. Shoots of plants grown under PT conditions are shown in Figure S2.
(d) Primary roots of plants treated as described in (c). Roots of plants grown under PT conditions are shown in Figure S2.
(e) Coloring of GUS staining solution after incubation with mock and 5, 20 and 50 μM inhibitor-treated plants. Every experimental point included eight plants, which were incubated with 1 ml of the staining solution for 16 h. Note that the amount of GUS induced by DZNep DT (c,d) was not sufficient for cleavage of the substrate in free solution. Furthermore, the lower signal intensity of PT-treated samples treated with a high concentration of DZNep (50 μM) indicates strongly reduced plant viability.
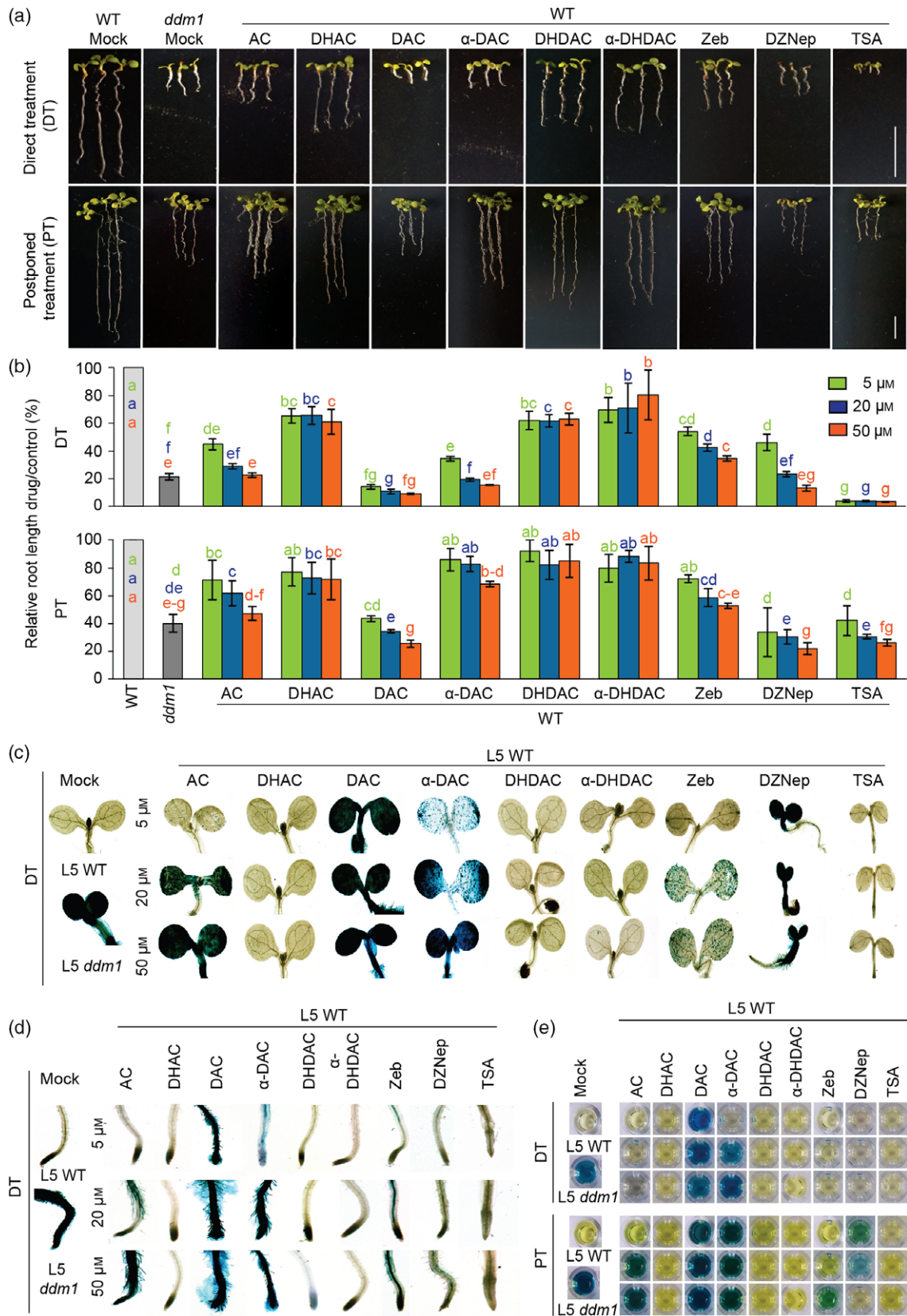
**Figure 3.** Drug-induced changes in DNA methylation.

(a) Immunolocalization of 5-methyl-2′-deoxycytidine (5-mdC; green) on Arabidopsis nuclei isolated from wild type (WT) plants without drug treatment (mock), treated with epigenetic drugs (20 μM) and *ddm1* plants serving as low DNA methylation control. Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (grey in merge). AC, 5-azacytidine; DAC, 2′-deoxy-5-azacytidine; Zeb, zebularine; DZNep, 3-deazaneplanocin A. Scale bar = 10 μm.

(b) High-precision liquid chromatography-based quantification of 5-mdCs in DNA of mock- and drug-treated wild-type and *ddm1* plants. Error bars denote standard deviations between the means of three biological replicates. Experimental points marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$).

(c) Methylation-sensitive Southern blot analysis of mock- and drug-treated wild-type (WT) plants and *ddm1* plants grown under the direct treatment (DT) and postponed treatment (PT) regimes. Genomic DNA was digested with *Hpa*II, *Msp*I and *Alu*I and hybridized with centromeric repeat (*pAL*) and 5S rDNA probes.

(d) Reverse transcription quantitative PCR on DNA methylation-controlled genetic elements *LINE1-4*, *TSI*, *TsGUS*, *Ta3*, *soloLTR* and *SDC*, and in seedlings treated with 0 (mock) and 40 μM drugs for 48 h. Error bars represent standard deviation between means of three biological replicates. The values were normalized to the *PP2A* gene. Means that do not share a letter are significantly different according to Tukey's test ($P \leq 0.05$).

Arabidopsis (Pecinka *et al.*, 2009; Liu *et al.*, 2015). Therefore, we compared the inhibitor-induced single-strand annealing type HR frequency using a disrupted GUS reporter gene system (Puchta *et al.*, 1995). Under the treatment with a low concentration (2.5 μM), we observed a significantly increased HR frequency for Zeb, while the other treatments did not differ from WT (AC, DAC) or were even reduced (DZNep) (Figure 6d). Therefore, we repeated the experiment with a higher concentration (10 μM). The results remained the same except for DAC, which significantly increased HR frequency.

An important parameter is inhibitor cytotoxicity. To estimate the amount of inhibitor-induced cell death, we performed propidium iodide (PI) staining on roots of living plants treated with 20 μM inhibitors for 24 or 48 h (Figure 6e). Mock-treated WT plants showed no cell death. 5-Azacytidine and Zeb induced cell death in the differentiated zone which was progressively spreading towards the meristematic zone. Interestingly, DAC-induced cell death was localized mainly in the meristematic zone. Finally, DZNep treatment caused cell death all over the root meristem and elongation zone.

## DISCUSSION

We performed comparative analyses of nine epigenetic inhibitors, representing three functionally distinct groups, with respect to their TGS suppressing and DNA damaging effects. The histone deacetylase (HDA) class I and II inhibitor TSA causes histone hyper-acetylation, changes in gene expression and developmental alterations (Xu *et al.*, 2005; Rosa *et al.*, 2014). Recently it was shown that some plants naturally release precursors of HDA inhibitors in order to suppress their competitors (Venturelli *et al.*, 2015). We confirmed high TSA toxicity but did not observe release of TGS from *TsGUS* locus. This was unexpected, because *TsGUS* is activated by a wide spectrum of mutations in genes controlling TGS, including *HISTONE DEACETYLASE 6* (*HDA6*) mutants, and the activation can occur even without a loss of DNA methylation (Elmayan *et al.*, 2005; Pecinka *et al.*, 2010). However, it corresponds with the finding that TSA and AC analogs often have opposite effects (Chang and Pikaard, 2005). We hypothesize that TSA represses multiple

essential HDAs necessary for the control of housekeeping genes including, for example, HDA19 (Tanaka *et al.*, 2008), which probably override the phenotypic effects of HDA6 repression (Aufsatz *et al.*, 2002).

3-Deazaneplanocin A is an inhibitor of SAHH, which produces methyl groups for methylation of proteins and DNA (Miranda *et al.*, 2009). Genetic loss of *SAHH* is lethal, and hypomorphic alleles show severe developmental phenotypes (Rocha *et al.*, 2005; Baubec *et al.*, 2010). 3-Deazaneplanocin A activates stably silenced transgenes (Baubec *et al.*, 2010; Foerster *et al.*, 2011). We show that DZNep strongly suppresses TGS in shoots, but not in roots. The molecular basis of this difference is unknown but could be caused by tissue-specific DZNep uptake, its metabolization or the need for methyl groups. 3-Deazaneplanocin A is highly cytotoxic and lethal for Arabidopsis at a concentration of 50 μM.

The most abundant group of tested chemicals were the AC analogs DAC and α-DAC, several 5,6-dihydro derivatives (DHAC, DHDAC, α-DHDAC) and Zeb. In human cell culture, 5,6-dihydro compounds induced a moderate to strong reduction in DNA methylation (Matoušová *et al.*, 2011). By contrast, previous data from *Nicotiana benthamiana* (Mynarzova and Baranek, 2015) and our data from Arabidopsis suggest that DHAC, DHDAC and α-DHDAC do not activate TGS-controlled loci in plants. At present, it is not clear whether this is due to problems in uptake, incorporation or stability of these compounds or whether fundamental differences exist between DNA methylation establishment and maintenance pathways in plants and mammals (reviewed in, e.g., Feng *et al.*, 2010). 5-Azacytidine and Zeb led to moderate activation and DAC with α-DAC to strong activation of the TGS target loci. The most effective inhibitor in almost all experiments was DAC. At the same time, it was the most cytotoxic cytidine analog. However, cytotoxicity of DAC was slightly weaker than that of DZNep and was similar for root and shoot, making DAC the preferred drug of choice for epigenetic studies.

We confirmed the DNA-demethylating potential for all core inhibitors (AC, DAC, Zeb, DZNep) on tandem repetitive sequences known to be controlled by TGS in
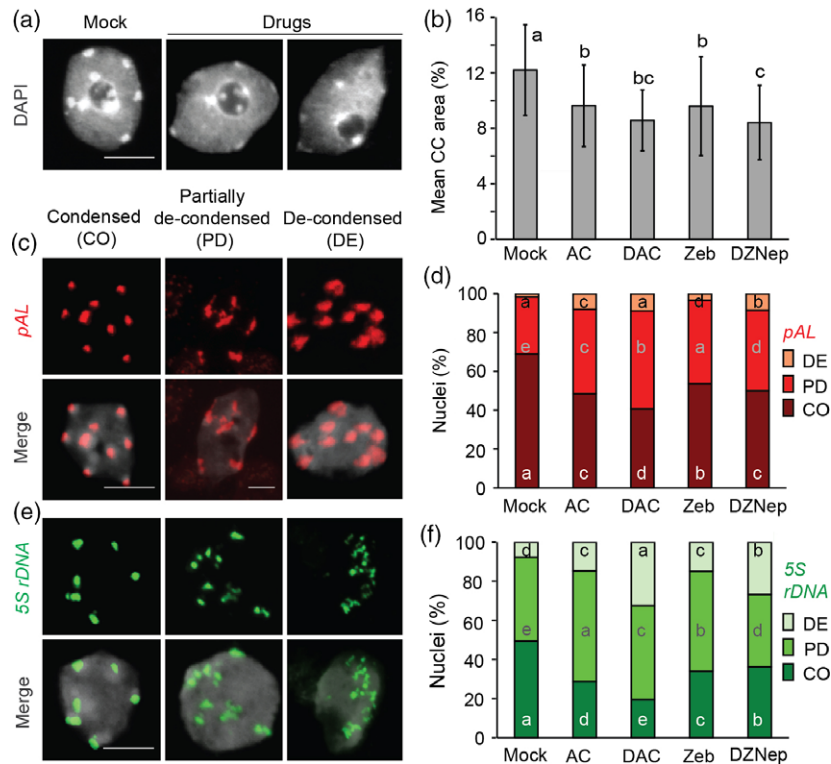
**Figure 4.** Effect of drugs on heterochromatin amount and chromocenter (CC) organization in Arabidopsis.

(a), (b) Heterochromatin quantification. (a) Representative 4′,6-diamidino-2-phenylindole (DAPI)-stained interphase nuclei isolated from control (mock) and 20 μM drug-treated wild-type Arabidopsis plants. Heterochromatic CCs appear as bright foci, euchromatin is grey and nucleoli are visible as a weakly stained spherical region. AC, 5-azacytidine; DAC, 2′-deoxy-5-azacytidine; Zeb, zebularine; DZNep, 3-deazaneplanocin A. Scale bar = 10 μm. (b) Mean CC area of nuclei prepared as described in (a). At least 100 nuclei per experimental point were analyzed. Error bars indicate the standard deviation between individual measurements. Values marked with the same letter do not differ according to Tukey's multiple range test ($P \leq 0.05$).

(c)–(e) Morphology of condensed (CO), partially de-condensed (PD) and de-condensed (DE) centromeric repeat (*pAL*, red, c) and *5S rDNA* (green, e) fluorescence *in situ* hybridization (FISH) signals. DNA was counterstained with DAPI (grey in merge). Scale bars = 10 μm. (d,f) Frequency of nuclei with CO, PD and DE FISH signals isolated from mock- and 20 μM drug-treated wild-type plants. At least 120 nuclei per experimental point were analyzed. Values marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$).

Arabidopsis, but we showed that it was dependent on the drug, treatment regime (DT versus PT), sequence context and also the locus tested. Direct treatment led to a generally higher degree of change in DNA methylation, which is consistent with our previous observation that cotyledon tissues are more responsive to such treatments due to suppressed RdDM (Baubec *et al.*, 2014). At the tested tandem repeats, DNA demethylation was observed mainly for the CG context, supporting that AC-type inhibitors affect DNMT1-like enzymes (Ghoshal *et al.*, 2005; Champion *et al.*, 2010), represented by MET1 in plants (Kankel *et al.*, 2003). In the CHG context, AC analogs caused minimal change at tandem repeats, but we observed demethylation by DZNep, which targets SAHH and thus biosynthesis of the whole methyl group. Hence, DZNep presumably reduces DNA methylation by removing the substrate for methylation reactions, rather than directly inhibiting CHG DNA methyltransferase CMT3. Finally, we found an increase in DNA methylation in the CHH context, which

was most prominent for 5S rDNA repeats. We speculate that the transcripts derived from repeats are processed by the canonical RdDM pathway into small RNAs used to slice the transcripts and/or to hyper-methylate and thus to re-silence the target locus. This is consistent with our earlier genetic experiment showing that RdDM restores silencing at drug-activated TGS target loci (Baubec *et al.*, 2014). Drug-induced DNA demethylation was always weaker than genetically induced DNA demethylation in *ddm1* mutant plants, and drugs had generally stronger plant growth-suppressive effects. It has to be emphasized that the effects of epigenetic inhibitors on DNA methylation at regions other than highly repetitive ones may differ. Genome-wide analysis of DNA methylation at single base-pair resolution detected non-uniformly reduced levels of DNA methylation after treatments with high doses (100 μM) of AC and Zeb (Griffin *et al.*, 2016). The heterochromatic regions showed generally stronger demethylation compared with euchromatic ones with few differences between DNA methylation
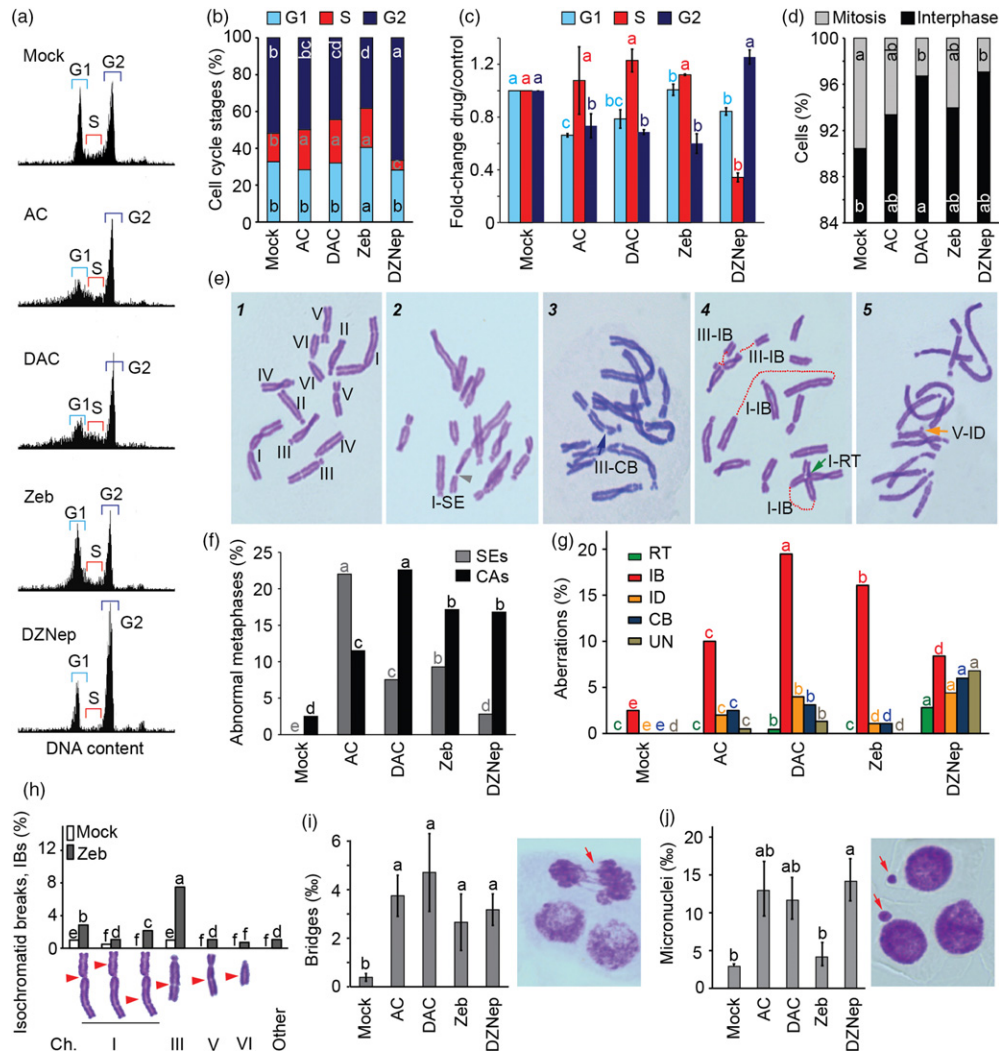
**Figure 5.** Analysis of inhibitor cytotoxicity in *Vicia faba*.

(a) Effect of drug treatment on the cell cycle in root apical meristems of *Vicia faba* plants. Four-day-old plants were treated with 20 μM drugs for 24 h and the nuclear DNA content of root apical meristem nuclei was measured by flow cytometry. The x-axis shows DNA content (relative fluorescence on $\log_3$ scale) and the y-axis the number of measured particles. Peaks correspond to nuclei in $G_1$ and $G_2$ stages. The space in between both peaks is considered to represent S-phase nuclei. AC, 5-azacytidine; DAC, 2′-deoxy-5-azacytidine; Zeb, zebularine; DZNep, 3-deazaneplanocin A.

(b) Quantification of nuclei in different stages of the cell cycle based on flow cytometric analysis described in (a). Each experimental point represents the mean of 10 independent measurements. Values marked with the same letter do not differ according to Tukey's multiple range test ($P \le 0.05$).

(c) Fold change of nuclei in different stages of the cell cycle relative to mock-treated wild-type based on flow cytometric analysis described in (a) and data from (b). Error bars represent standard deviations between ten samples. Values marked with the same letter do not differ according to Tukey's test ($P \le 0.05$).

(d) Percentage of mitotic (grey) and interphase (black) cells in squashes of root apical meristems of mock and 20 μM drug-treated *V. faba* plants. Every experimental point represents the mean of eight independent measurements (slides), each with at least 700 counted cells. Values marked with the same letter do not differ according to Tukey's test ($P \le 0.05$).

(e) Analysis of chromosome aberrations in metaphase plates of the faba bean karyotype ACB using the classification of (Fučík *et al.*, 1970). The plants were treated as described in (a,b): (1) typical metaphase plate in the mock-treated plants containing six individually distinguishable chromosome pairs (roman numbers); (2) example of segment extensions (SEs) on the short arm of chromosome I (grey arrowhead) after AC treatment; (3) the lower chromosome III carries a chromatid break (CB; dark blue arrow); (4) three isochromatid breaks (IB; broken fragments belonging to one chromosome are connected by red lines), and a symmetric reciprocal chromatid translocation (RT; green arrow) between the short and the long arm of one chromosome I; (5) interstitial deletion (ID, orange arrow) at chromosome V after DAC treatment. The dot-like deleted fragment remains attached to the intact sister chromatid.

(f) The frequency of segment extensions (SE) and chromatid structural aberrations (CA) in metaphase chromosomes. Values marked with the same letter do not differ according to Tukey's test ($P \le 0.05$). Numerical data are presented in Table S2.

(g) The frequency of CAs: RT, reciprocal translocation; IB, isochromatid break; ID, interstitial deletion; CB, chromatid breaks; UN, unclear cases with many aberrations. Values marked with the same letter do not differ according to Tukey's test ($P \le 0.05$). Numerical data are presented in Table S2.

(h) Frequency and typical chromosomal locations (red arrowheads) of IBs after 24 h of Zeb treatment. Ch., chromosome. Numerical data are presented in Table S3. Values marked with the same letter do not differ according to Tukey's test ($P \le 0.05$).

(i), (j) Frequency of root apical meristem nuclei with (i) anaphase bridges and (j) micronuclei (red arrows) per 1000 cells (‰). Every experimental point represents the mean of eight independent measurements (slides), each with at least 700 counted cells. Values marked with the same letter do not differ according to Tukey's test ($P \le 0.05$).
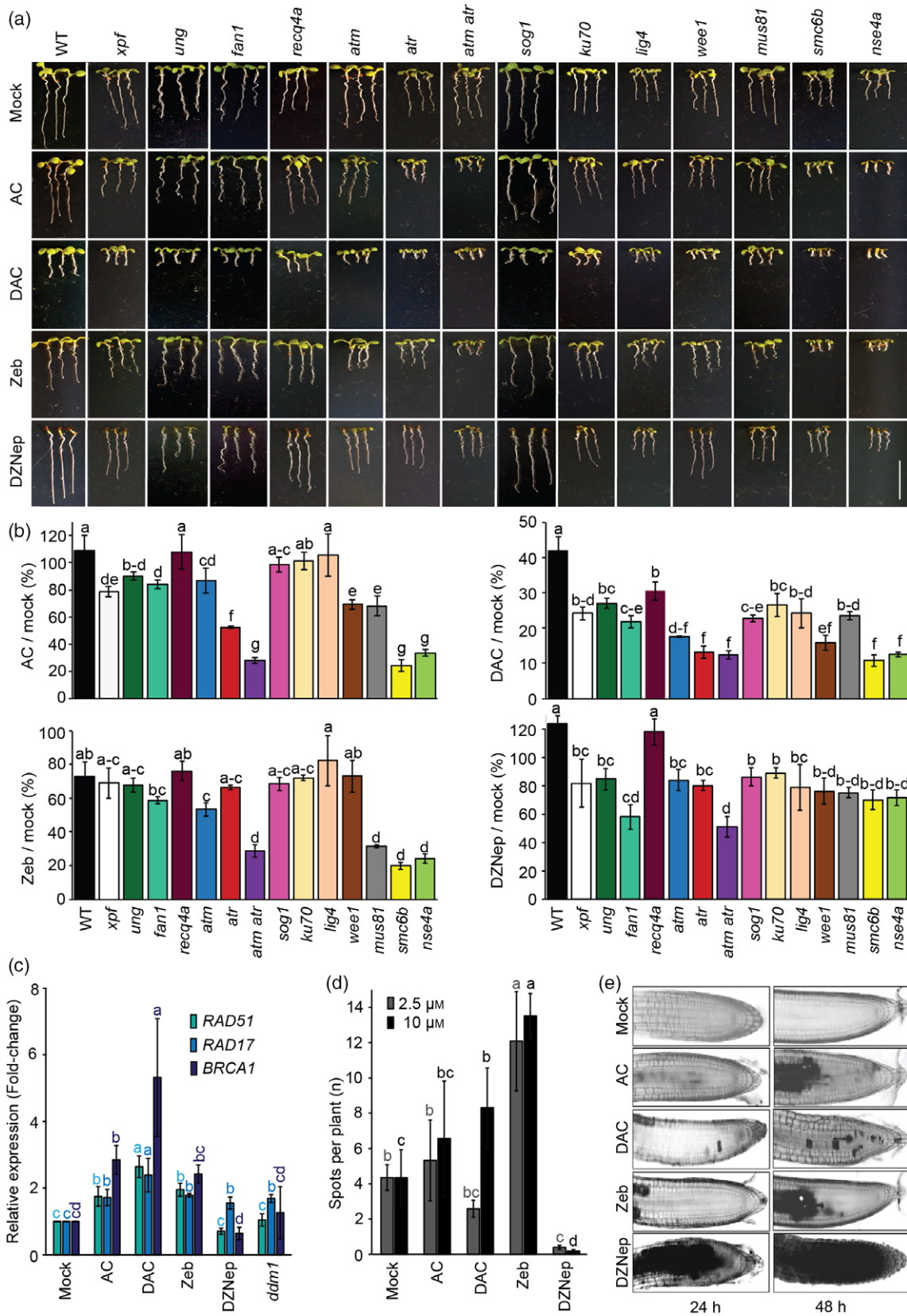
(a)

(b)

(c)

(d)

(e)

**Figure 6.** The sensitivity of Arabidopsis DNA damage repair (DDR) mutants to epigenetic inhibitors.
(a) Representative growth phenotypes of wild type (WT) and DDR mutant plants in response to 5 μM concentrations of epigenetic inhibitors under the direct treatment regime. AC, 5-azacytidine; DAC, 2′-deoxy-5-azacytidine; Zeb, zebularine; DZNep, 3-deazaneplanocin A. Scale bar = 10 mm.
(b) Quantitative data for (a) calculated as the relative root length of one genotype under drug/mock conditions. Error bars represent the standard deviation between means of three biological replicates. The replicates were grown on separate screening plates and each contained at least 25 plants. Values marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$).
(c) Reverse transcription quantitative PCR analysis of DDR marker genes *RAD51*, *RAD17* and *BRCA1* in seedlings treated with 0 (mock) and 20 μM concentration of drugs for 24 h. Error bars represent the standard deviation between the means of four biological replicates. The values were normalized to the *PP2A* gene. Experimental points marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$).
(d) Inhibitory effect on homologous recombination using single-strand annealing reporter line B11 in response to 2.5 and 10 μM concentrations of chemicals expressed as number of GUS spots per plant. Error bars represent mean of three biological replicates, each with at least 30 plants. Values marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$).
(e) Analysis of cell death after inhibitor treatment. Propidium iodide stained roots from living WT seedlings treated as mock and by 20 μM concentration of drugs for 24 and 48 h. Dark regions indicate dead cells.

---

contexts. Although it is currently unknown whether this is due to a weaker DNA demethylation potential or faster re-methylation in euchromatin, it shows that the inhibitor effects are locus-specific. It is also clear that the tested epigenetic inhibitors cannot fully substitute genetic mutations but can be a useful tool for studying the impact of epigenetic changes in species where mutants are not easily available and/or a transient effect is needed.

To add another facet to the picture of the effects of epigenetic inhibitors on plants, we analyzed their impact on genome stability and DDR. Multiple experiments, including analysis of the cell cycle, frequency of cell division, chromosome structural aberrations and micronuclei, suggested that AC, DAC, Zeb and DZNep reduce genomic stability. The hallmark of AC, DAC and Zeb treatments were segment extensions (i.e. strong decondensation of specific heterochromatic regions of metaphase chromosomes) and isochromatid breaks. Aberrations induced by DZNep treatment were more diverse, indirectly suggesting that the damage caused by AC analogs and DZNep might be different. But the actual nature of this damage is currently unknown. Our previous and current work, based on the comet assays and the candidate gene screens, suggests formation of single-strand (but not double-strand) breaks after AC treatment as the immediate damage (Liu *et al.*, 2015). Data from bacteria, fungi and animals suggest that AC-like compounds covalently bind DNA methyltransferases, which leads to the formation of bulky adducts and the accumulation of specific bubble-, X- and Y-shaped DNA structures (Kuo *et al.*, 2007; Salem *et al.*, 2009; Champion *et al.*, 2010). Hence, the damage could be represented by DNA methyltransferase–cytidine analog-specific DNA–protein crosslinks (reviewed in Stingele *et al.*, 2017), but the evidence for such a type of damage is missing in plants.

Candidate gene screening revealed that DNA damage induced by AC, DAC, Zeb and DZNep is signaled by ATM and ATR kinases. While ATR is the major kinase signaling AC- and DAC-induced damage, both ATM and ATR signal the presence of Zeb- and DZNep-induced lesions. Nucleotide excision repair, BER and NHEJ seem to represent only minor pathways for the repair of inhibitor-induced damage. By contrast, experiments with the mutants and HR traps indicate that HR is the preferred pathway for repair of inhibitor-induced damage and the repair of damage caused by AC analogs is highly dependent on the SMC5/6 complex. The mechanism of DDR by the SMC5/6 complex remains unknown, but this complex has been shown to be important for detoxifying aberrant X-shaped intermediates that occur during DNA replication in *Saccharomyces cerevisiae* (Menolfi *et al.*, 2015) and to promote HR by enhancing sister chromatid alignment in Arabidopsis (Watanabe *et al.*, 2009). Since AC was shown to produce aberrant plasmid structures in *Escherichia coli* (Kuo *et al.*, 2007), it is possible that the high sensitivity of Arabidopsis SMC5/6 mutants results from accumulation of such toxic structures or from reduced sister chromatid alignment, which is required for correct homology-dependent DNA repair. However, experimental testing of these hypotheses is difficult, because Arabidopsis meristems are very small and methods for their efficient cell-cycle synchronization are not yet available.

By combining information from both epigenetic and genome stability parts of our study, we conclude that there is a strong correlation between the degree of TGS repression and induction of DNA damage by cytidine-like compounds (AC, DAC, Zeb). This suggests that the formation of DNA damage may be an integral part of the cytidine analog-induced DNA demethylation. Speculatively, demethylation occurs via depletion of the DNMT1 pool, as observed in mammals (Ghoshal *et al.*, 2005), in combination with ongoing DNA replication and/or strand synthesis during DDR, which would lead to the synthesis of stretches of hemi- or un-methylated DNA. However, drug-induced demethylation does not last for a long time in meristematic plant cells due to high activity of epigenetic factors, including the *de novo* DNA methyltransferase DRM2 (Baubec *et al.*, 2014). Here, the most effective drug in terms of reducing TGS was DAC, but at the same time this drug is associated with high cytotoxicity, possibly due to its direct incorporation into DNA without the need for metabolization and time-dependent degradation.

## EXPERIMENTAL PROCEDURES

### Plant materials, growth conditions and chemical treatments

We used *A. thaliana* wild-type Columbia-0 (Col-0), L5 reporter line (Morel *et al.*, 2000; Elmayan *et al.*, 2005), B11 (N1DC1 no. 11) line containing an intramolecular type of HR substrate (Puchta *et al.*, 1995) and the following mutants: *mus81-1* (GK_113F11), *ku70* (SALK_123114C), *lig4-2* (SALK_044027C), *recq4a-4* (GK_203C07), *smc6b-1* (SALK_101968), *nse4a-2* (GK-768H08), *xpf-3* (SALK_096156C), *wee1-1* (GK_270E05), *ung-1* (GK-440E07), *sog1-1* (EMS mutant G155R; Yoshiyama *et al.*, 2009), *fan1-1* (GK_815C08), *atm-2* (SALK_006953, −/− were selected from the segregating F$_2$ population), *atr-2* (SALK_032841C) and double homozygous *atm-2 atr-2* (selected from progeny of +/− and −/− plants, respectively) in the Col-0 background. *ddm1-5* L5 was in the Col/Zh background (Mittelsten Scheid *et al.*, 1998) and Col-0 background, respectively. The *V. faba* used contained the ACB karyotype (Döbel *et al.*, 1978).

*Arabidopsis thaliana* seeds were surface-sterilized as described (Baubec *et al.*, 2009), stratified in the dark at 4°C for 48 h, evenly spread on solid half-strength Murashige–Skoog (½MS) medium with 0.6% agar and grown under 16-h light/8-h dark cycles at 21°C as indicated in Figure 1(b) in the absence or presence of 5, 20 and 50 μM inhibitors. All inhibitors were synthesized at the Institute of Organic Chemistry and Biochemistry, Prague, as described (Matoušová *et al.*, 2011), except for DZNep and TSA, which were purchased commercially (Sigma-Aldrich, https://www.sigmaaldrich.com/; SML0305 and T8552, respectively). For heterochromatin quantification, FISH and immunostaining, seedlings were grown on plates containing ½MS solid medium for 7 days and then transferred to plates containing ½MS medium with 0 (mock) and 20 μM concentrations of drugs for 3 days. For HR analysis, plants were grown for 10 days on solid ½MS medium supplemented with 0, 2.5 or 10 μM of drugs. For cell death analysis, seedlings were grown for 3 days on solid ½MS medium prior to transfer for 24 or 48 h to liquid ½MS medium without or with 20 μM of the drug. For quantitative PCR, seedlings were grown on plates containing ½MS solid medium for 7 days and then transferred for 24 h to liquid ½MS medium without or with 20 μM of drugs. At some experimental points, the seedlings were transferred for 48 h to liquid ½MS medium without (mock) or 40 μM of inhibitors (see the text).

*Vicia faba* seeds were germinated for 3 days on a wet filter paper at room temperature (21°C) in the dark. Primary roots about 1–2 cm long were incubated for 24 h in aerated Hoagland solution. For chemical treatments, the roots were incubated for 24 h in distilled water containing 0 and 20 μM drug, followed by incubation in Hoagland solution for 4.5 h. For studies of chromatid aberration, root tips were in addition exposed to 0.05% colchicine for 2.5 h (to arrest metaphase) and then fixed in ethanol:glacial acetic acid (3:1) overnight.

### Root length assay, GUS staining and cell death analysis

For root length assay, plants were carefully pulled out of the medium using forceps without breaking the primary roots and then stretched on agar plates. Seedlings were photographed with a D90 digital camera (Nikon, https://www.nikon.com/) and the roots were measured using ImageJ calibrated with an internal size control. Sensitivity to each chemical treatment in individual replicates was determined by calculating mean(treatment)/mean (mock). The roots of at least 25 plants per treatment were measured for each of the three biological replicates. The GUS histochemical staining was performed as described (Baubec *et al.*, 2009). All GUS samples were analyzed and photographed using a SZX16 binocular microscope equipped with Regita 1300 QImaging camera and QCAPTURE x64 software (both Olympus, https://www.olympus-global.com/). For cell death assay, seedlings were stained with 10 μg ml$^{-1}$ of PI solution (Sigma) for 3 min, followed by a rinsing step with sterilized water, placed on slides in a drop of water and then evaluated using the AxioImager Z2 (Zeiss, https://www.zeiss.com/) microscope equipped with a high performance DSD2 confocal module (Andor, https://andor.oxinst.com/).

### DNA isolation and DNA methylation assays

Genomic DNA was extracted from entire seedlings using Nucleon PhytoPure Kit (GE Healthcare, https://www.gehealthcare.com/) with additional RNase I digestion. Total cytosine methylation was determined by high-precision liquid chromatography as described (Finke *et al.*, 2018). All samples were analyzed in triplicate, and 5mdC values were expressed as a percentage of total cytosine. The probes for Southern hybridization specific for centromeric repeat (*pAL*) and *5S rDNA* were prepared from *A. thaliana* Col-0 genomic DNA and directly labeled using PCR with biotin-dUTP (Roche, https://www.roche.com/). For Southern blot analysis, 350 ng of genomic DNA was digested overnight at 37°C with 1 unit of *Hpa*II, *Msp*I or *Alu*I (all Thermo Scientific, https://www.thermofisher.com/). Subsequently, the samples were electrophoretically separated overnight on 1.2% 2-amino-2-(hydroxymethyl)-1,3-propanediol (TRIS)–borate–EDTA agarose gels, depurinated, denatured and neutralized as described (Baubec *et al.*, 2009). Gels were blotted onto Amersham Hybond N+ (https://www.gelifesciences.com/) membranes for 7–8 h with 20× SSC, washed in 2× SSC, dried for 30 min and UV-crosslinked using a STRATALINKER (Agilent, https://www.agilent.com/en/products/genomics-agilent). Hybridization was performed as described (Southern, 1975) with modifications. In brief, hybridization buffer contained 5× SSC, 5× Denhardt´s solution and 2% SDS. For pre-hybridization, the membrane was first incubated for 60 min at 68°C in 20 ml of hybridization buffer and later for another 60 min under the same conditions with the hybridization buffer supplemented with 1 mg of salmon sperm DNA (Invitrogen; AM9680). Before adding the salmon sperm DNA to pre-hybridization buffer, 100 μl of it was boiled for 10 min and cooled on ice for 10 min (final concentration of salmon sperm DNA = 50 ng ml$^{-1}$). For hybridization, approximately 100 ng of probe was prepared as above to a final concentration of 5 ng ml$^{-1}$ and hybridized overnight at 68°C. The next day, membranes were washed for 5 min each in Wash-1 (2× SSC, 0.1% SDS) and 5 min in Wash-2 (0.2× SSC, 0.1% SDS) buffers at room temperature, followed by two washes (15 min each) in Wash-2 (0.2× SSC, 0.1% SDS) and one wash for 15 min in Wash-3 (0.1× SSC, 0.1% SDS) buffers at 54°C for the *pAL* probe and 61°C for the *5S rDNA* probe. Biotin-labeled *pAL* (centromeric repeat) and *5S rDNA* probes were detected using Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific, 89880) following the manufacturer's directions.

### RNA isolation and quantitative PCR

Total RNA was isolated using the RNeasy Plant Mini kit (Qiagen, https://www.qiagen.com/). Complementary DNA was synthesized from 1 μg of total RNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, https://www.thermofisher.com/) with oligo dT primers according to the manufacturer's instructions. The cDNA was diluted 1:5 and RT-qPCR was performed using 2 μl of cDNA per 20 μl reaction with the 5× HOT FIREPol Eva Green qPCR Mix Plus (ROX) kit (Solis Biodyne, https://www.solisbiodyne.com/) on an CFX96 Touch Real-Time PCR Detection System (Bio-Rad, https://www.bio-rad.com/). Fold

changes were calculated relative to a mock-treated control using the standard curve method. Quantitative PCR experiments were performed following the MIQE guidelines (Bustin *et al.*, 2009). *PP2A* (AT1G69960) was used as the reference gene. The primers used in this study are listed in Table S3.

## Cytology

Nuclei were isolated by chopping whole seedlings with a razor blade in isolation buffer (100 mM TRIS pH = 7.0, 50 mM KCl, 2 mM $MgCl_2$, 0.05% Tween 20, 5% sucrose) and filtering through a 30-μm nylon mesh. The suspension of nuclei was centrifuged to microscopic slides using Cytospin (MPW Medical Instruments, https://mpw.pl/) as described (Baubec *et al.*, 2009). Preparations were air-dried, post-fixed in 4% formaldehyde in PBS and stored at −20°C until use. The FISH probes specific for the centromeric repeat (*pAL*) and *5S rDNA* were amplified from *A. thaliana* Col-0 genomic DNA and directly labeled with biotin-dUTP and digoxigenin-dUTP (Roche) respectively, during PCR (Probst *et al.*, 2003). Slide pre-treatment, hybridization, post-hybridization washes and detection steps were carried out as described (Pecinka *et al.*, 2004, 2010). Biotin-dUTP was detected by goat anti-avidin conjugated with biotin (1:200; Vector Laboratories, https://vectorlabs.com/) and avidin combined with Texas-Red (1:1000; Vector Laboratories), digoxigenin-dUTP by mouse anti-digoxigenin (1:250; Roche) and goat anti-mouse conjugated with Alexa[488] (1:200; Thermo Fisher Scientific). Immunolocalization of methylated cytosine was performed as described (Baubec *et al.*, 2009). Slides were incubated with the primary monoclonal mouse-anti-5-methylcytosine (1:200; Diagonode, https://www.diagenode.com/) and the secondary goat anti-mouse-Alexa[488] (1:200; Thermo Fisher Scientific) antibodies. Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI; 1 μg ml$^{-1}$) in Vectashield (Vector Laboratories). All images were captured using an Axiocam 503 monochromatic camera attached to Axio Imager.A2 epifluorescence microscope (both Zeiss). Images were captured separately for each fluorochrome using the appropriate excitation and emission filters with ZEN (Zeiss) system. The monochromatic images were pseudocolored and merged using Adobe Photoshop CS5 (Adobe Systems, https://www.adobe.com/) software. Digital images in grayscale were analyzed with IMAGEJ using custom-made plugins (Data S1 and S2). The scripts were written to measure the size and average staining intensity of nuclei and CCs. The CC value was divided by the whole nucleus value to yield the CC fraction.

## Cell cycle, cell division and chromatid structural analysis using *V. faba*

For cell cycle analysis by flow cytometry, faba bean root tips were dissected, chopped with a razor blade in 300 ml of extraction buffer (Sysmex, https://www.sysmex.co.jp), filtered through 30-μm nylon mesh, stained with 900–1800 μl CyStain dye (Sysmex) and analyzed with a PAS I ploidy analyzer (Sysmex). Ten individual plants per treatment were measured. For analysis of cell division and chromatid structural changes, roots were washed for 10 min in distilled water, hydrolyzed for 11 min in 1 N HCl at 60°C, stained for 30–40 min in Feulgen solution and squashed in a drop of 45% acetic acid. Microscopic analysis was performed with an Axio Imager.A2 epifluorescence microscope, an Axiocam 503 mono camera and ZEN system (all Zeiss) or an ECLIPSE-E600 modular microscope, equipped with a DS-RI1 camera and NIS-Elements system (all Nikon). Eight to twelve slides, each corresponding to one RAM of one plant, were evaluated per experimental point. In total 700 cells per slide were scored for quantification of mitotic divisions, anaphase bridges and micro-nuclei. The same

preparations were used for observations of chromatid structural changes. At least 200 complete metaphase cells from five slides were scored to quantify segment extensions and chromatid aberrations. Segment extension regions were evaluated as described (Fučík *et al.*, 1970). Chromatid aberrations were classified as follows: chromatid and isochromatid breaks (one or both sister chromatids with terminal deletion), interstitial deletions, duplication deletion (the deleted part of one chromatid is inserted into a break of the sister chromatid) and reciprocal chromatid translocations.

## Statistical analysis

The values were examined by one-way analysis of variance (ANOVA) and post-hoc Tukey's honestly significant difference (HSD) test ($P \leq 0.05$) using STATISTICA v. 13 (StatSoft, http://www.statsoft.com/) or MINITAB v. 18 (Minitab, LLC, https://www.minitab.com/) programs.

## CONFLICT OF INTEREST

The authors declare no competing financial interests.

## AUTHOR CONTRIBUTIONS

AP and AN designed the research. AN, BT, JZ, EDT, KP, UE, AF, WR, BP and IS performed experiments. MO and MK synthesized azacytidine and its analogs. IN wrote the plugin to measure nuclear and CC areas in IMAGEJ software. AN analyzed the data and prepared the figures. AP and AN wrote the manuscript with contributions from all authors. All authors approved the final version of this article.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1**. Plant growth in response to inhibitor treatment.

**Figure S2**. L5 reporter locus activation in response to inhibitor treatment.

**Figure S3**. Drug-induced DNA demethylation changes.

**Table S1**. Frequency of chromosome aberrations in metaphase plates of the field bean karyotype ACB after 24 h of drug treatment of root tip meristems.

**Table S2**. Frequency of isochromatic breaks recognized in metaphase plates of the field bean karyotype ACB after 24 h of zebularine treatment of root tip meristems.

**Table S3**. Primers used in this study.

**Data S1**. Plugin to measure nuclear area in IMAGEJ software.

**Data S2**. Plugin to measure chromocenter area in IMAGEJ software.

# REFERENCES

**Alberts, B.** (2002) Chromosomal DNA and its packaging in the chromatin fiber. In *Molecular Biology of the Cell* (Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. eds). New York: Garland science.

**van Attikum, H., Bundock, P., Overmeer, R.M., Lee, L., Gelvin, S.B. and Hooykaas, P.J.J.** (2003) The Arabidopsis AtLIG4 gene is required for the repair of DNA damage, but not for the integration of Agrobacterium T-DNA. *Nucleic Acids Res.* **31**, 4247–4255.

**Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M. and Matzke, A.J.M.** (2002) HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J.* **21**, 6832–6841.

**Baubec, T., Pecinka, A., Rozhon, W. and Mittelsten Scheid, O.** (2009) Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *Plant J.* **57**, 542–554.

**Baubec, T., Dinh, H.Q., Pecinka, A., Rakic, B., Rozhon, W., Wohlrab, B., von Haeseler, A. and Scheid, O.M.** (2010) Cooperation of multiple chromatin modifications can generate unanticipated stability of epigenetic states in Arabidopsis. *Plant Cell*, **22**, 34–47.

**Baubec, T., Finke, A., Mittelsten Scheid, O. and Pecinka, A.** (2014) Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Rep.* **15**, 446–452.

**Bewick, A.J., Ji, L. and Niederhuth, C.E.** et al. (2016) On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl Acad. Sci. USA*, **113**, 9111–9116.

**Bustin, S.A., Benes, V., Garson, J.A.** et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**: 611–622.

**Champion, C., Guianvarc'h, D., Sénamaud-Beaufort, C., Jurkowska, R.Z., Jeltsch, A., Ponger, L., Arimondo, P.B. and Guieysse-Peugeot, A.-L.** (2010) Mechanistic insights on the inhibition of C5 DNA methyltransferases by zebularine. *PLoS ONE*, **5**, e12388.

**Chang, S. and Pikaard, C.S.** (2005) Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation. *J. Biol. Chem.* **280**, 796–804.

**Cho, S.-W., Ishii, T., Matsumoto, N., Tanaka, H., Eltayeb, A.E. and Tsujimoto, H.** (2011) Effects of the cytidine analogue zebularine on wheat mitotic chromosomes. *Chromosom. Sci.* **14**, 23–28.

**Córdoba-Cañero, D., Dubois, E., Ariza, R.R., Doutriaux, M.-P. and Roldán-Arjona, T.** (2010) Arabidopsis Uracil DNA glycosylase (UNG) is required for base excision repair of uracil and increases plant sensitivity to 5-fluorouracil. *J. Biol. Chem.* **285**, 7475–7483.

**Culligan, K.M., Robertson, C.E., Foreman, J., Doerner, P. and Britt, A.B.** (2006) ATR and ATM play both distinct and additive roles in response to ionizing radiation. *Plant J.* **48**, 947–961.

**De Schutter, K., Joubès, J., Cools, T.** et al. (2007) Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. *Plant Cell*, **19**, 211–225.

**Diaz, M., Pecinkova, P., Nowicka, A., Baroux, C., Sakamoto, T., Yuliani Gandha, P., Jeřábková, H., Matsunaga, S., Grossniklaus, U. and Pecinka, A.** (2019) SMC5/6 complex subunit NSE4A is involved in DNA damage repair and seed development in Arabidopsis. *Plant Cell*, **31**, 1579–1597.

**Döbel, P., Schubert, I. and Rieger, R.** (1978) Distribution of heterochromatin in a reconstructed karyotype of Vicia faba as identified by banding- and DNA-late replication patterns. *Chromosoma*, **69**, 193–209.

**Du, J., Zhong, S. and Bernatavichute, Y.V.** et al. (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167–180.

**Elmayan, T., Proux, F. and Vaucheret, H.** (2005) Arabidopsis RPA2: A genetic link among transcriptional gene silencing, DNA repair, and DNA replication. *Curr. Biol.* **15**, 1919–1925.

**Fajkus, J., Vyskot, B. and Bezděk, M.** (1992) Changes in chromatin structure due to hypomethylation induced with 5-azacytidine or DL-ethionine. *FEBS Lett.* **314**, 13–16.

**Fauser, F., Schiml, S. and Puchta, H.** (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* **79**, 348–359.

**Feng, S., Jacobsen, S.E. and Reik, W.** (2010) Epigenetic reprogramming in plant and animal development. *Science*, **330**, 622–627.

**Fidantsef, A.L., Mitchell, D.L. and Britt, A.B.** (2000) The Arabidopsis UVH1 gene is a homolog of the yeast repair endonuclease RAD1. *Plant Physiol.* **124**, 579–586.

**Finke, A., Rozhon, W. and Pecinka, A.** (2018) Analysis of DNA methylation content and patterns in plants. *Methods Mol. Biol.* **1694**, 277–298.

**Fiskus, W., Wang, Y. and Sreekumar, A.** et al. (2009) Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells. *Blood*, **114**, 2733–2743.

**Foerster, A.M., Dinh, H.Q., Sedman, L., Wohlrab, B. and Mittelsten Scheid, O.** (2011) Genetic rearrangements can modify chromatin features at epialleles. *PLoS Genet.* **7**, e1002331.

**Fojtová, M., Kovařík, A., Votruba, I. and Holý, A.** (1998) Evaluation of the impact of S-adenosylhomocysteine metabolic pools on cytosine methylation of the tobacco genome. *Eur. J. Biochem.* **252**, 347–352.

**Fransz, P., Soppe, W. and Schubert, I.** (2003) Heterochromatin in interphase nuclei of *Arabidopsis thaliana*. *Chromosom. Res.* **11**, 227–240.

**Fučík, V., Michaelis, A. and Rieger, R.** (1970) On the induction of segment extension and chromatid structural changes in Vicia faba chromosomes after treatment with 5-azacytidine and 5-azadeoxycytidine. *Mutat. Res. Mol. Mech. Mutagen.* **9**, 599–606.

**Fulneček, J., Matyášek, R., Votruba, I., Holý, A., Křížová, K. and Kovařík, A.** (2011) Inhibition of SAH-hydrolase activity during seed germination leads to deregulation of flowering genes and altered flower morphology in tobacco. *Mol. Genet. Genomics*, **285**, 225–236.

**Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T. and Jacob, S.T.** (2005) 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. *Mol. Cell. Biol.* **25**, 4727–4741.

**Glazer, R.I., Knode, M.C., Tseng, C.K.H., Haines, D.R. and Marquez, V.E.** (1986) 3-deazaneplanocin A: a new inhibitor of S-adenosylhomocysteine synthesis and its effects in human colon carcinoma cells. *Biochem. Pharmacol.* **35**, 4523–4527.

**Griffin, P.T., Niederhuth, C.E. and Schmitz, R.J.** (2016). A comparative analysis of 5-azacytidine and zebularine induced DNA demethylation. *G3: Genes - Genomes - Genetics*, **6**, 2773–2780.

**Hegde, V., McFarlane, R.J., Taylor, E.M. and Price, C.** (1996) The genetics of the repair of 5-azacytidine-mediated DNA damage in the fission yeast Schizosaccharomyces pombe. *Mol. Gen. Genet.* **251**, 483–492.

**Herrmann, N.J., Knoll, A. and Puchta, H.** (2015) The nuclease FAN1 is involved in DNA crosslink repair in *Arabidopsis thaliana* independently of the nuclease MUS81. *Nucleic Acids Res.* **43**, 3653–3666.

**Hu, Z., Cools, T. and De Veylder, L.** (2016) Mechanisms used by plants to cope with DNA damage. *Annu. Rev. Plant Biol.* **67**, 439–462.

**Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A.J.M. and Matzke, M.** (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J.* **25**, 2828–2836.

**Jackson, J.P., Lindroth, A.M., Cao, X. and Jacobsen, S.E.** (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature*, **416**, 556–560.

**Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L. and Richards, E.J.** (2003) Arabidopsis MET1 cytosine Methyltransferase mutants. *Genetics*, **163**, 1109–1122.

**Kawakatsu, T., Hwang, S.C. and Jupe, F.,** et al. (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell*, **166**, 492–505.

**Kimura, S. and Sakaguchi, K.** (2006) DNA repair in plants. *Chem. Rev.* **106**, 753–766.

**Kiziltepe, T., Hideshima, T. and Catley, L.** et al. (2007) 5-Azacytidine, a DNA methyltransferase inhibitor, induces ATR-mediated DNA double-strand break responses, apoptosis, and synergistic cytotoxicity with doxorubicin and bortezomib against multiple myeloma cells. *Mol. Cancer Ther.* **6**, 1718–1727.

**Kovarik, A., Van Houdt, H., Holý, A. and Depicker, A.** (2000) Drug-induced hypomethylation of a posttranscriptionally silenced transgene locus of tobacco leads to partial release of silencing. *FEBS Lett.* **467**, 47–51.

**Kuo, H.K., Griffith, J.D. and Kreuzer, K.N.** (2007) 5-azacytidine–induced Methyltransferase-DNA adducts block DNA replication in vivo. *Cancer Res.* **67**, 8248–8254.

Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, **133**, 523–536.

Liu, C. and Weigel, D. (2015) Chromatin in 3D: progress and prospects for plants. *Genome Biol.* **16**, 170.

Liu, C.-H., Finke, A., Diaz, M., Rozhon, W., Poppenberger, B., Baubec, T. and Pecinka, A. (2015) Repair of DNA damage induced by the cytidine analog zebularine requires ATR and ATM in Arabidopsis. *Plant Cell*, **27**, 1788–1800.

Lyko, F. and Brown, R. (2005) DNA Methyltransferase inhibitors and the development of epigenetic cancer therapies. *J. Nat. Cancer Inst.*, **97**, 1498–1506.

Mathieu, O., Reinders, J., Čaikovski, M., Smathajitt, C. and Paszkowski, J. (2007) Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell*, **130**, 851–862.

Matoušová, M., Votruba, I., Otmar, M., Tloušťová, E., Günterová, J. and Mertlíková-Kaiserová, H. (2011) 2′-deoxy-5,6-dihydro-5-azacytidine—a less toxic alternative of 2′-deoxy-5-azacytidine: a comparative study of hypomethylating potential. *Epigenetics*, **6**, 769–776.

Matzke, M.A. and Mosher, R.A. (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408.

Meier, I., Richards, E.J. and Evans, D.E. (2017) Cell biology of the plant nucleus. *Annu. Rev. Plant Biol.* **68**, 139–172.

Mengiste, T., Revenkova, E., Bechtold, N. and Paszkowski, J. (1999) An SMC-like protein is required for efficient homologous recombination in Arabidopsis. *EMBO J.* **18**, 4505–4512.

Menolfi, D., Delamarre, A., Lengronne, A., Pasero, P. and Branzei, D. (2015) Essential roles of the Smc5/6 complex in replication through natural pausing sites and endogenous DNA damage tolerance. *Mol. Cell*, **60**, 835–846.

Miranda, T.B., Cortez, C.C., Yoo, C.B., Liang, G., Abe, M., Kelly, T.K., Marquez, V.E. and Jones, P.A. (2009) DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol. Cancer Ther.* **8**, 1579–1588.

Mittelsten Scheid, O., Afsar, K. and Paszkowski, J. (1998) Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. *Proc. Natl Acad. Sci. USA*, **95**, 632–637.

Moissiard, G., Cokus, S.J. and Cary, J. *et al.* (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. *Science*, **336**, 1448–1451.

Morel, J.-B., Mourrain, P., Béclin, C. and Vaucheret, H. (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr. Biol.* **10**, 1591–1594.

Mynarzova, Z. and Baranek, M. (2015) An evaluation of the impact of demethylation agents treatment using TGS 16C *Nicotiana benthamiana* reporter line. *MendelNet*, 428–433.

Nowicka, A., Juzoń, K., Krzewska, M., Dziurka, M., Dubas, E., Kopeć, P., Zieliński, K. and Żur, I. (2019) Chemically-induced DNA de-methylation alters the effectiveness of microspore embryogenesis in triticale. *Plant Sci.* **287**, 110189.

O'Malley, R.C., Barragan, C.C. and Ecker, J.R. (2015) A user's guide to the Arabidopsis T-DNA insertional mutant collections. *Methods Mol. Biol. (Clifton, N.J.)*, **1284**, 323–342.

Orta, M.L., Höglund, A., Calderón-Montaño, J.M., Domínguez, I., Burgos-Morón, E., Visnes, T., Pastor, N., Ström, C., López-lázaro, M. and Helleday, T. (2014) The PARP inhibitor Olaparib disrupts base excision repair of 5-aza-2′-deoxycytidine lesions. *Nucleic Acids Res.* **42**, 9108–9120.

Ossowski, S., Schwab, R. and Weigel, D. (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* **53**, 674–690.

Pecinka, A. and Liu, C.-H. (2014) Drugs for plant chromosome and chromatin research. *Cytogenet Genome Res.* **143**, 51–59.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M., Fuchs, J. and Schubert, I. (2004) Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. *Chromosoma*, **113**, 258–269.

Pecinka, A., Rosa, M., Schikora, A., Berlinger, M., Hirt, H., Luschnig, C. and Scheid, O.M. (2009) Transgenerational stress memory is not a general response in Arabidopsis. *PLoS ONE*, **4**, e5202.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N. and Scheid, O.M. (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell*, **22**, 3118–3129.

Probst, A.V., Fransz, P.F., Paszkowski, J. and Scheid, O.M. (2003) Two means of transcriptional reactivation within heterochromatin. *Plant J.* **33**, 743–749.

Puchta, H., Swoboda, P. and Hohn, B. (1995) Induction of intrachromosomal homologous recombination in whole plants. *Plant J.* **7**, 203–210.

Riha, K., Watson, J.M., Parkey, J. and Shippen, D.E. (2002) Telomere length deregulation and enhanced sensitivity to genotoxic stress in Arabidopsis mutants deficient in Ku70. *EMBO J.* **21**, 2819–2826.

Rocha, P.S.C.F., Sheikh, M., Melchiorre, R., Fagard, M., Boutet, S., Loach, R., Moffatt, B., Wagner, C., Vaucheret, H. and Furner, I. (2005) The Arabidopsis HOMOLOGY-DEPENDENT GENE SILENCING1 gene codes for an S-adenosyl-l-homocysteine hydrolase required for DNA methylation-dependent gene silencing. *Plant Cell*, **17**, 404–417.

Rosa, S., Ntoukakis, V., Ohmido, N., Pendle, A., Abranches, R. and Shaw, P. (2014) Cell differentiation and development in Arabidopsis are associated with changes in histone dynamics at the single-cell level. *Plant Cell*, **26**, 4821–4833.

Salem, A.M.H., Nakano, T., Takuwa, M., Matoba, N., Tsuboi, T., Terato, H., Yamamoto, K., Yamada, M., Nohmi, T. and Ide, H. (2009) Genetic analysis of repair and damage tolerance mechanisms for DNA-protein cross-links in *Escherichia coli*. *J. Bacteriol.* **191**, 5657–5668.

Schiml, S., Fauser, F. and Puchta, H. (2016) Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. *Proc. Natl Acad. Sci. USA*, **113**, 7266–7271.

Solís, M.-T., El-Tantawy, A.-A., Cano, V., Risueño, M.C. and Testillano, P.S. (2015) 5-azacytidine promotes microspore embryogenesis initiation by decreasing global DNA methylation, but prevents subsequent embryo development in rapeseed and barley. *Front. Plant Sci.* **6**, 472.

Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I. and Fransz, P.F. (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J.* **21**, 6549–6559.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517.

Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Scheid, O.M. and Paszkowski, J. (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. *Plant Cell*, **12**, 1165–1178.

Stingele, J., Bellelli, R. and Boulton, S.J. (2017) Mechanisms of DNA–protein crosslink repair. *Nat. Rev. Mol. Cell Biol.* **18**, 563.

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V. and Jacobsen, S.E. (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*, **152**, 352–364.

Takuno, S. and Gaut, B.S. (2012) Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Mol. Biol. Evol.* **29**, 219–227.

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.

Venturelli, S., Belz, R.G. and Kämper, A., *et al.* (2015) Plants release precursors of Histone Deacetylase inhibitors to suppress growth of competitors. *Plant Cell*, **27**, 3175–3189.

Watanabe, K., Pacher, M., Dukowic, S., Schubert, V., Puchta, H. and Schubert, I. (2009) The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 complex promotes sister chromatid alignment and homologous recombination after DNA damage in *Arabidopsis thaliana*. *Plant Cell*, **21**, 2688–2699.

Xu, C.-R., Liu, C., Wang, Y.-L., Li, L.-C., Chen, W.-Q., Xu, Z.-H. and Bai, S.-N. (2005) Histone acetylation affects expression of cellular patterning genes in the Arabidopsis root epidermis. *Proc. Natl Acad. Sci. USA*, **102**, 14469–14474.

Yoo, C.B. and Jones, P.A. (2006) Epigenetic therapy of cancer: past, present and future. *Nat. Rev. Drug Discov.* **5**, 37–50.

Yoshiyama, K., Conklin, P.A., Huefner, N.D. and Britt, A.B. (2009) Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc. Natl Acad. Sci. USA*, **106**, 12843–12848.

**Zadražil, S., Fučík, V., Bartl, P., Šormová, Z. and Šorm, F.** (1965) The structure of DNA from *Escherichia coli* cultured in the presence of 5-azacytidine. *Biochim. Biophys. Acta – Nucleic Acids Protein Synth.* **108**, 701–703.

**Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L. and Zilberman, D.** (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, **153**, 193–205.

**Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. and Henikoff, S.** (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**, 61–69.

**Publikace 3**

Check for updates

# The SMC5/6 Complex Subunit NSE4A Is Involved in DNA Damage Repair and Seed Development[OPEN]

Mariana Díaz,[a,b,c] Petra Pečinková,[a] Anna Nowicka,[a,b,d] Célia Baroux,[e] Takuya Sakamoto,[c] Priscilla Yuliani Gandha,[a] Hana Jeřábková,[b] Sachihiro Matsunaga,[c] Ueli Grossniklaus,[e] and Ales Pecinka[a,b,1]

[a] Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research (MPIPZ), 50829 Cologne, Germany
[b] The Czech Academy of Sciences, Institute of Experimental Botany (IEB), Centre of the Region Haná for Biotechnological and Agricultural Research, 77900 Olomouc, Czech Republic
[c] Department of Applied Biological Science, Faculty of Science and Technology, Tokyo University of Science, Chiba 278-8510, Japan
[d] Polish Academy of Sciences, Franciszek Gorski Institute of Plant Physiology, 30-239 Krakow, Poland
[e] Department of Plant and Microbial Biology and Zürich-Basel Plant Science Center, University of Zürich, 8008 Zürich, Switzerland

ORCID IDs: 0000-0003-3887-3897 (M.D.); 0000-0001-8275-4620 (P.P.); 0000-0002-5762-3482 (A.N.); 0000-0001-6307-2229 (C.B.); 0000-0002-8896-5944 (T.S.); 0000-0001-5500-6184 (P.Y.G.); 0000-0003-3748-8997 (H.J.); 0000-0003-3024-3559 (S.M.); 0000-0002-0522-8974 (U.G.); 0000-0001-9277-1766 (A.P.)

The maintenance of genome integrity over cell divisions is critical for plant development and the correct transmission of genetic information to the progeny. A key factor involved in this process is the STRUCTURAL MAINTENANCE OF CHROMOSOME5 (SMC5) and SMC6 (SMC5/6) complex, related to the cohesin and condensin complexes that control sister chromatid alignment and chromosome condensation, respectively. Here, we characterize *NON-SMC ELEMENT4* (*NSE4*) paralogs of the SMC5/6 complex in Arabidopsis (*Arabidopsis thaliana*). NSE4A is expressed in meristems and accumulates during DNA damage repair. Partial loss-of-function *nse4a* mutants are viable but hypersensitive to DNA damage induced by zebularine. In addition, *nse4a* mutants produce abnormal seeds, with noncellularized endosperm and embryos that maximally develop to the heart or torpedo stage. This phenotype resembles the defects in cohesin and condensin mutants and suggests a role for all three SMC complexes in differentiation during seed development. By contrast, *NSE4B* is expressed in only a few cell types, and loss-of-function mutants do not have any obvious abnormal phenotype. In summary, our study shows that the NSE4A subunit of the SMC5-SMC6 complex is essential for DNA damage repair in somatic tissues and plays a role in plant reproduction.

## INTRODUCTION

The eukaryotic nuclear genome is packaged into higher order chromatin structures that are dynamically remodeled during cellular activities (Alabert and Groth, 2012). Key factors establishing and orchestrating chromosome organization are STRUCTURAL MAINTENANCE OF CHROMOSOME (SMC) complexes: cohesin (containing SMC1 and SMC3), condensin (containing SMC2 and SMC4), and the SMC5/6 complex (containing SMC5 and SMC6; reviewed in Hirano, 2006; Jeppsson et al., 2014b; Uhlmann, 2016). The heterodimeric SMC backbone serves as a structural component and a docking platform for additional subunits that vary depending on the complex, thereby enabling a variety of specific assemblies (reviewed in Kegel and Sjögren, 2010; Diaz and Pecinka, 2018). Studies in yeasts and animals showed that cohesin facilitates sister chromatin cohesion, and condensin I and II complexes mediate large-scale chromatin folding and chromosome condensation (reviewed in Hirano, 2012; Uhlmann, 2016). The major activity of the SMC5/6 complex is the maintenance of nuclear genome stability by resolving complex structures and possibly acting as an antagonist of the cohesin complex (reviewed in De Piccoli et al., 2009; Kegel and Sjögren, 2010; Diaz and Pecinka, 2018). The SMC5/6 complex performs many functions, such as the control of unidirectional rDNA replication, neutralizing toxic DNA intermediates during replication, preventing homologous recombination between nonhomologous sequences, and alternative telomere lengthening (Potts and Yu, 2007; Torres-Rosell et al., 2007; Chiolo et al., 2011; Menolfi et al., 2015).

The SMC5/6 complex can be associated with up to six NON-STRUCTURAL ELEMENT (NSE) subunits, which assemble in a combinatorial manner to form three subcomplexes (NSE1-NSE3-NSE4, NSE5-NSE6, and NSE2-SMC5-SMC6) in yeasts (De Piccoli et al., 2009; Duan et al., 2009). Studies in budding yeast, fission yeast, and mammalian cell cultures revealed that the NSE1-NSE3-NSE4 subcomplex binds double-stranded DNA and acts as a binding platform for the heads of SMC5 and SMC6 (Hudson et al., 2011; Palecek and Gruber, 2015; Zabrady et al., 2016. The least evolutionary conserved SMC5/6 complex subunits are NSE5 and NSE6. They interact with the SMC5-SMC6 hinges in budding yeast but with their heads in fission yeast (Pebernard et al., 2006; De Piccoli et al., 2009; Duan et al., 2009). Recently, functional orthologs of NSE5 and NSE6 have been identified in plants and mammals (Yan et al., 2013; Räschle et al., 2015), but their molecular functions remain unclear. NSE2 (also

## IN A NUTSHELL

**Background:** The nuclear genome is organized into chromosomes, which are dynamically remodelled during cellular activities. The STRUCTURAL MAINTENANCE OF CHROMOSOME (SMC) complexes are key factors establishing and orchestrating chromosome organization. Among the SMC complexes are cohesin (facilitating sister chromatid cohesion), condensin (mediating large-scale chromosome folding), and SMC5/6 (maintaining genome stability). The SMC5/6 complex is composed of eight subunits that form its ring structure and perform specific functions. Existing data show that this SMC5/6 complex has a number of unexpected roles in plants including control of specific developmental processes or suppression of hyper-immune responses.

**Question:** We wanted to describe functions of two sister genes for NSE4 of Arabidopsis, which remained as the last fully uncharacterized subunit of SMC5/6 complex.

**Findings:** We found that *NSE4A* and *NSE4B* genes originate from a whole-genome duplication about 30 million years ago. NSE4A is expressed in somatic and reproductive tissue and accumulates upon induction of DNA damage. Complete loss of function of NSE4A is lethal, but we found a partial-loss-of-function mutant that is sensitive to specific types of DNA damage. Moreover, the mutant shows poor development of seeds, where the embryo and its surrounding nutritive tissue endosperm stop developing at early stages. By contrast, the function of NSE4B remains unclear because it is expressed only in few tissues and plants lacking a functional *NSE4B* gene look normal. By genetic modification and expression of NSE4B, we show that it is not efficient in DNA damage repair.

**Next steps:** How the SMC5/6 complex controls genome stability and affects plant growth is still not understood. We are deciphering this by analysing phenotypes of multiple mutants in the SMC5/6 complex using mixture of genetic, molecular and biochemical methods.

known as METHANE METHYLSULFONATE SENSITIVE21 [MMS21] and HIGH PLOIDY2 [HPY2]) is anchored to SMC5 and has SMALL UBIQUITIN-RELATED MODIFIER E3 ligase activity (Zhao and Blobel, 2005). Many proteins were found to be targets of NSE2 sumoylation, including several SMC5/6 and cohesin subunits, as well as DNA repair proteins in plants, fungi, and animals (Zhao and Blobel, 2005; Pebernard et al., 2006; Potts and Yu, 2007; Huang et al., 2009; Ishida et al., 2009).

Homologs of all SMC5/6 complex subunits were identified in Arabidopsis (*Arabidopsis thaliana*; Schubert, 2009; Watanabe et al., 2009; Yan et al., 2013; Diaz and Pecinka, 2018). However, our understanding of biological processes controlled by the individual SMC5/6 complex subunits remains limited in plants. Arabidopsis plants mutated in *SMC6B* (also known as *HYPER-SENSITIVE TO MMS*, *IRRADIATION*, *AND MITOMYCIN C* [*MMC*]) are indistinguishable from the wild type under ambient conditions but are hypersensitive to DNA damaging treatments, show a delayed repair of DNA strand breaks, and have a reduced frequency of homologous recombination (Mengiste et al., 1999; Kozak et al., 2009; Watanabe et al., 2009; Liu et al., 2015). *smc6a* mutants are viable even under severe DNA damage, but *smc6a smc6b* double mutation is embryo lethal (Watanabe et al., 2009; Yan et al., 2013), indicating partial functional redundancy. Plants defective in *NSE2* are hypersensitive to DNA damage and display a wide range of pleiotropic phenotypes, including leaf and stem malformations, branching defects, reduced meristem size, impaired development of gametes, shortened vegetative phase, and increased drought tolerance (Huang et al., 2009; Ishida et al., 2009; Xu et al., 2013; Zhang et al., 2013; Liu et al., 2014; Yuan et al., 2014; Kwak et al., 2016). SMC5, SMC6, and NSE1, NSE2, NSE3 and NSE4 are evolutionary conserved proteins. In addition, there are two other SMC5/6 complex subunits (collectively named as NSE5 and NSE6) in fungi, animals, and plants, which are presumably functionally conserved but share little sequence similarity (reviewed in Diaz and Pecinka, 2018). In Arabidopsis, both the

regulator of systemic acquired resistance SUPPRESSOR OF NPR1-1, INDUCIBLE1 (SNI1) and the ARABIDOPSIS SNI1 ASSOCIATED PROTEIN1 (ASAP1) were found in a complex with SMC5 and SMC6B and were thus proposed as the putative functional orthologs of yeast *NSE6* and *NSE5*, respectively (Yan et al., 2013). Both genes participate in the control of genome stability and suppression of immune hyper-responses, which is a novel and unexpected function of the complex.

The variety of plant phenotypes seen in mutants affecting the SMC5/6 complex suggests that it participates in multiple developmental and cellular pathways possibly linked to stress responses. Currently, the composition of the plant SMC5/6 complex, the roles of individual subunits, and their functional requirement in cellular and developmental processes (besides DNA damage repair) are poorly characterized. In an effort to obtain a more comprehensive functional understanding of the Arabidopsis SMC5/6 complex, we characterized the roles of the NSE4A and NSE4B subunits. We show that *NSE4A* is involved in repair of zebularine-induced DNA damage in challenged somatic tissues. In addition, *NSE4A* is essential for reproductive development in Arabidopsis, while the function of *NSE4B* remains elusive.

## RESULTS

### The *NSE4* Gene Is Duplicated in the Arabidopsis Genome

The Arabidopsis genome contains two uncharacterized, putative, *NSE4* homologs: *NSE4A* (At1g51130 encoding a 403 amino acid protein) and *NSE4B* (At3g20760 encoding a 383 amino acid protein) sharing 65.1% identity at the amino acid level (Figures 1A and 1B). To identify the age of this duplication, we built a *NSE4* phylogeny across green plants using the *Schizosaccharomyces pombe* and *Homo sapiens* NSE4s as outgroups (Figure 1C; Supplemental Table 1; Supplemental Data Sets 1 to 3). Except
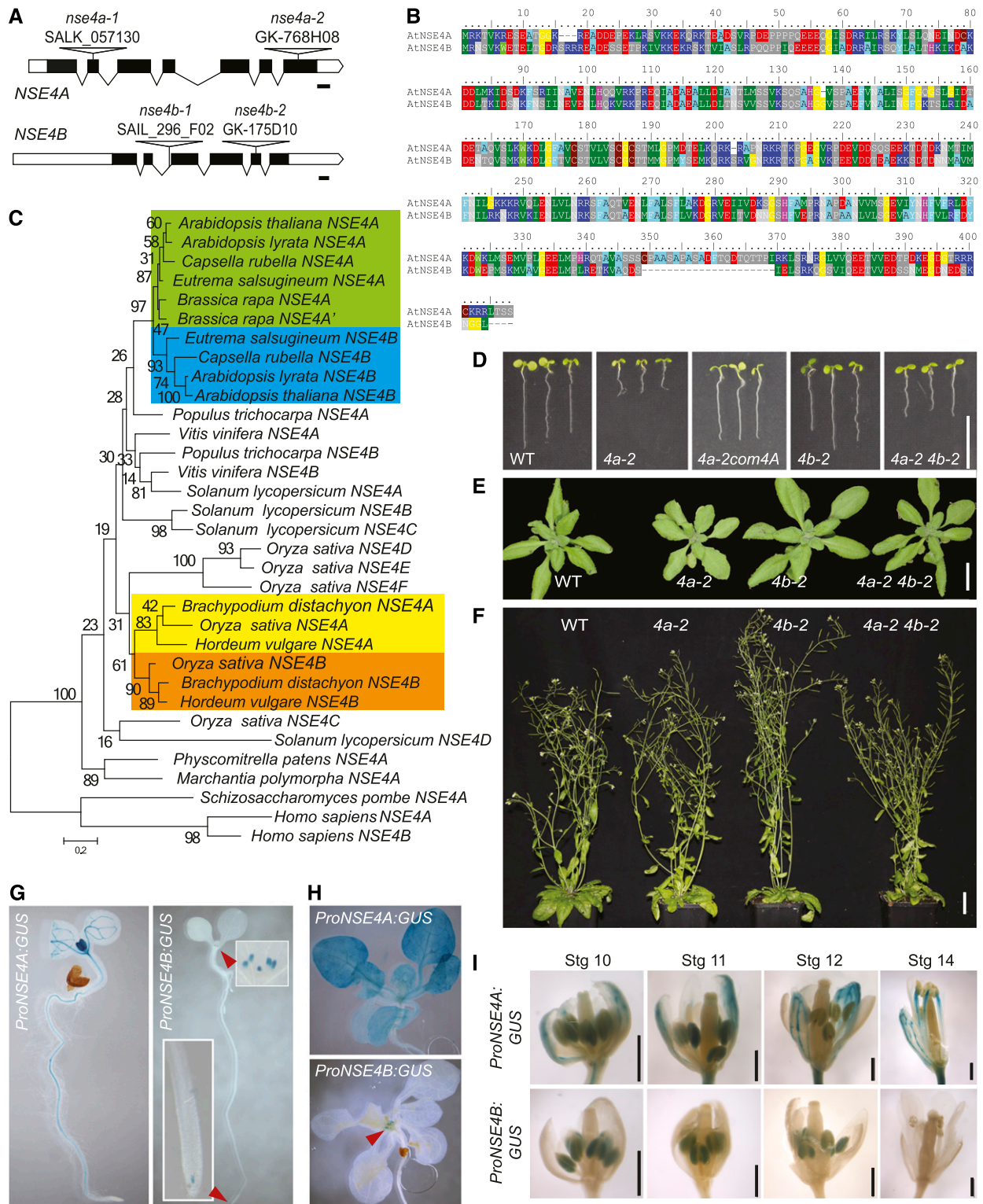
**Figure 1.** Basic Characterization of *NSE4* Paralogs.

**(A)** Gene structure of *A. thaliana NSE4A* and *NSE4B* with indicated positions of the mutations used in this study. Bars, 100 bp.

**(B)** Alignment of Arabidopsis NSE4A and NSE4B proteins.

forBryophyta and Marchantiophyta, which carry a single *NSE4*, all other plant genomes contained at least two *NSE4* copies. Orthologs of Arabidopsis *NSE4A* and *NSE4B* occurred in *Arabidopsis lyrata*, *Capsella rubella*, and *Eutrema salsugineum*. The only exception was *Brassica rapa*, where both *NSE4* copies were derived from *NSE4A*, while *NSE4B* was missing. This suggests that the *NSE4A* and *NSE4B* originate from the whole-genome duplication event that occurred ~47 million years ago (MYA) and preceded radiation of the species within Brassicaceae (Kagale et al., 2014). Phylogenetic shadowing of *NSE4A* and *NSE4B* promoters revealed that both contain conserved blocks, A1 and B1, respectively, directly upstream of the transcription start site (Supplemental Figure 1; Supplemental Table 2). However, the A1 block was clearly larger and more similar between species, indicating that it may contain key *NSE4A* cis-regulatory sequences. There was another set of conserved *NSE4* paralogs in Poaceae, including *Brachypodium distachyon*, *Hordeum vulgare*, and *Zea mays* (Figure 1C). These paralogs most likely appeared during the Poaceae-specific whole-genome duplication event ~70 MYA (Paterson et al., 2009). We found a total of six *NSE4* copies in rice and four in tomato. Some of these copies were short and grouped with more distantly related species (Figure 1C), raising questions on their origin and functionality. The high frequency of multiple *NSE4* copies per genome may indicate rapid *NSE4* sub- or neo-functionalization in different plant lineages.

To assess the role of the *NSE4* genes in plant growth and development, we isolated T-DNA insertion mutations in *NSE4A* and *NSE4B* (Figure 1A). The *nse4a-1* allele carried a T-DNA in the second exon and was lethal as indicated by the absence of homozygous mutants in the progeny of heterozygous parents. However, we recovered viable homozygous *nse4a-2* plants carrying a T-DNA insertion in the last exon, 56 amino acids before the stop codon (Supplemental Figure 2). A 3′ rapid amplification of cDNA ends (RACE) technique revealed that the *NSE4A* transcript in *nse4a-2* plants continued into the T-DNA and maintained the reading frame for 201 nucleotides, adding a predicted 67 alien amino acids to the NSE4A protein produced by *nse4a-2* mutants. Therefore, *nse4a-2* most likely represents a partial loss-of-function allele with a modified C terminus. Juvenile and nonflowering *nse4a-2* plants were smaller than the wild type (Figures 1D and 1E) but recovered and were indistinguishable from control plants at flowering (Figure 1F). The *nse4b* mutant alleles carried T-DNA insertions in the second intron (*nse4b-1*) and the fifth exon (*nse4b-2*), respectively. Amplification from cDNA with primer pairs positioned on either side of the T-DNA insertions yielded very low or no products in

quantitative PCR, suggesting that both insertions disrupt the *NSE4B* transcript (Supplemental Figure 3). However, both *nse4b-1* and *nse4b-2* plants were viable and resembled the wild-type plants (Figures 1D to 1F). Combining the *nse4a-2* and *nse4b-2* alleles in a homozygous double mutant resulted in a *nse4a-2*–like phenotype, suggesting that *NSE4A* and *NSE4B* do not act redundantly during vegetative development.

To reveal the activity pattern of the *NSE4* promoter, we generated stable reporter lines where the *NSE4A* and *NSE4B* promoters were fused to the *uidA* gene encoding β-glucuronidase (*GUS*; *ProNSE4A:GUS* and *ProNSE4B:GUS*). The *NSE4A* promoter was strongly active in emerging true leaves and weakly active in the vasculature of the cotyledons at 7 d after germination (DAG; Figure 1G). In addition, we observed signals in the stele tissues within the differentiation zone of the root, but there was no *ProNSE4A* activity in root meristems. At 14 DAG, *ProNSE4A* was weakly active in all aerial tissues (Figure 1H). Flowers showed *ProNSE4A:GUS* activity in sepals, the upper half of fully elongated anther filaments, pistils, and anthers (Figure 1I, top). By contrast, *ProNSE4B:GUS* activity was restricted to the leaf stipules and a small domain in the root apical meristem at 7 DAG (Figure 1G, red arrowheads and insets). This pattern remained unchanged during the entire vegetative phase (Figure 1H). In flowers, *ProNSE4B* was active in anthers between stages 10 and 12 (Figure 1I). The difference in the expression patterns of *NSE4A* and *NSE4B* could be due to the association of the endogenous *NSE4B* locus with repressive histone H3 Lys-27 trimethylation (Supplemental Figures 4 and 5).

### *NSE4A* Is Expressed in Pollen, Ovules, and Seeds

The activity of *ProNSE4A* and *ProNSE4B* in flowers prompted us to analyze the reproductive stages in more detail. To get better insight into the expression of the NSE4A protein, we expressed a translational fusion of NSE4A with VENUS (an improved variant of the yellow fluorescent protein; Nagai et al., 2002) under the control of its native promoter (*ProNSE4A:NSE4A-VENUS*) in the *nse4a-2* background. Based on the full complementation of *nse4a-2* hypersensitivity to zebularine (Figure 2A), we conclude that the addition of VENUS does not interfere with NSE4A function.

Analysis of the transcription during pollen development revealed strong and weak activity of *ProNSE4A* and *ProNSE4B*, respectively (Figures 2B and 2C). The microspores (flower stage 10; Bowman et al., 1994) showed, on average, the strongest signals for both *ProNSE4A:GUS* and *ProNSE4B:GUS*, which decreased over subsequent developmental stages. There was

---

**Figure 1.** (continued).

**(C)** Phylogenetic tree of *NSE4* homologs in plants based on the maximum likelihood algorithm (see "Methods"). Fission yeast *NSE4/RAD62* and human *NSE4* paralogs were used as outgroups. Brassicaceae and Poaceae *NSE4* duplications are indicated by the colored squares. Identifiers of the protein sequences used to build the tree are provided as Supplemental Data Set 1.

**(D)** to **(F)** Phenotypes of the homozygous wild-type (WT), *nse4a-2* (*4a-2*), *nse4a-2* complemented with *ProNSE4A:GenomicNSE4A* (*4a-2 com4A*), *nse4b-2* (*4b-2*), and *nse4a-2 nse4b-2* (*4a-2 4b-2*) plants. **(D)** One-week-old in vitro–grown seedlings. Bar = 10 mm. **(E)** Three-week-old plants in soil. Bar = 25 mm. **(F)** Six-week-old mature plants. Bar = 35 mm.

**(G)** to **(I)** Analysis of *NSE4A* and *NSE4B* promoter activity using the GUS reporter system. **(G)** One-week-old plants grown as described in **(D)**. Red arrowheads indicate *ProNSE4B:GUS* signals in the root meristematic zone and leaf stipules (top inset). **(H)** Fourteen-day-old plants grown in in vitro culture. **(I)** Flowers at developmental stage (Stg) 10 to 14 (Bowman et al., 1994). Bars = 500 μm.
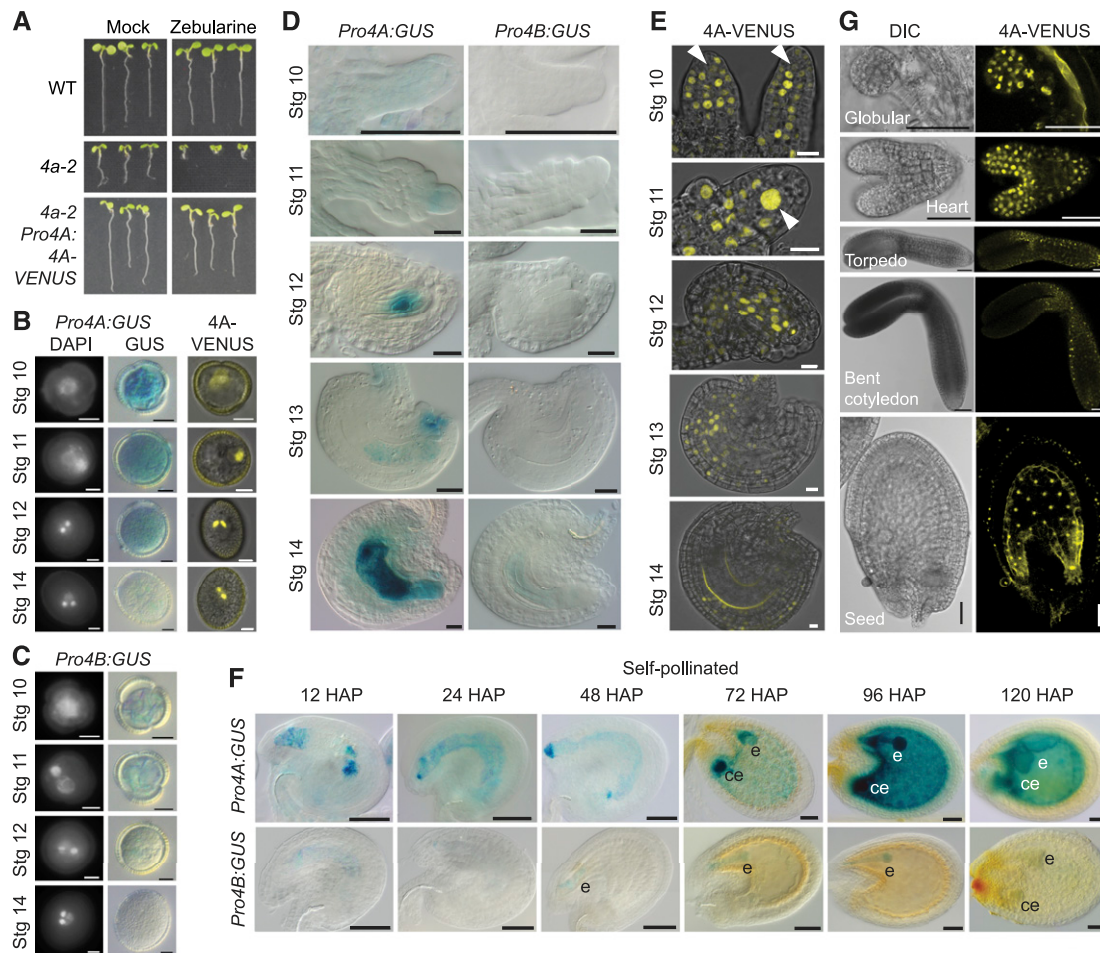
**Figure 2.** NSE4 Expression Analysis during Pollen, Ovule, and Seed Development.

**(A)** Test for functionality of NSE4-VENUS translational fusion line. Wild-type (WT), *nse4a-2* (*4a-2*), and *nse4a-2* plants complemented with *ProNSE4A*:*NSE4A*:*VENUS* (4A-VENUS) were germinated and grown on the control and zebularine-containing media for 7 d. Restoration of root growth in *4a-2 NSE4A-VENUS* indicates full functionality of the translational fusion protein.

**(B)** The first two columns show DAPI- and GUS-stained pollen of *ProNSE4A*:*GUS* (*Pro4A*:*GUS*) reporter line. Stage (Stg) 10 corresponds to the microspore, Stg 11 to bicellular pollen, Stg 12 to tricellular pollen, and Stg 14 to mature pollen from open anthers. The last column shows pollen from the *ProNSE4A*:*NSE4A*:*VENUS* (4A-VENUS) reporter line. Bar = 5 μm.

**(C)** The *ProNSE4B*:*GUS* (*Pro4B*:*GUS*) reporter line presented in the same way as in **(A)**. Bar = 5 μm. Stg, stage.

**(D)** GUS activity of *ProNSE4A*:*GUS* (*Pro4A*:*GUS*; left) and *ProNSE4B*:*GUS* (*Pro4B*:*GUS*; right) from ovule primordia to early postfertilization. Stage (Stg) 10, 11, and 12 to 14 show ovule primordia, the nucellus, and developing the embryo sac, respectively. Bars = 50 μm.

**(E)** *ProNSE4A*:*NSE4A*:*VENUS* (4A-VENUS) signals at the same stages as described in **(C)**. In the ovule primordia of stage (Stg) 11, the megaspore mother cell is almost free of 4A-VENUS signal (arrowheads). However, its expression is greatly increased in the female meiocyte of Stg 11 (arrowhead). Bar = 10 μm.

**(F)** GUS activity driven by the *NSE4A* and *NSE4B* promoters at the indicated hours after pollination (HAP). Reporter lines were pollinated with their own pollen 48 h after emasculation. Bars = 50 μm. e, embryo; ce, chalazal endosperm.

**(G)** Accumulation of *ProNSE4A*:*NSE4A*:*VENUS* (4A-VENUS) in nuclei of globular-, heart-, torpedo-, and bent cotyledon–stage embryos and syncytial endosperm 72 h after pollination. Left images represent differential interference contrast (DIC), and the right images show the VENUS signal. Bars = 50 μm.

practically no transcriptional activity of both genes in mature pollen from open anthers (flower stage 14). At the protein level, NSE4A was present at all pollen stages in the cell lineage leading to the sperm cells, as indicated by VENUS signals in the single nucleus of the unicellular microspore, the generative nucleus of bicellular pollen (flower stage 11), and the two sperm nuclei of tricellular pollen (Figure 2B). No NSE4A-VENUS signal could be observed in the vegetative nucleus.

During ovule development (Figure 2D), we observed *ProNSE4A*:*GUS* activity in ovule primordia at flower stage 10, the nucellus at stage 11, and the embryo sac in stages 12 to 14. The transcriptional profile was largely in agreement with NSE4A protein accumulation (Figure 2E). Strong NSE4A-VENUS signals were observed in almost all cells of the nucellus except for the megaspore mother cell, where the fusion protein was barely detectable (Figure 2E, flower stage 10, arrowhead). However,

NSE4A-VENUS accumulated strongly in female meiocytes initiating meiotic prophase I (Figure 2E, flower stage 11, arrowhead). The differences between GUS and VENUS signals could be due to different stability of GUS mRNA and/or protein compared with NSE4A-VENUS transcript and/or protein. After pollination, *ProNSE4A* activity was detected in the embryo and the chalazal endosperm and later (at 96 h after pollination) also in the syncytial endosperm (Figure 2F). This corresponds well with the strong NSE4A-VENUS signals in developing embryos (Figure 2G) and also the prominent localization to the nuclei of the syncytial endosperm (Figure 2G). By contrast, *ProNSE4B* activity during early ovule development remained largely below detection limit (Figure 2D), and we detected weak activity only in mature embryo sacs, with GUS activity getting stronger after pollination, leading to a clear signal in the early embryo up to the globular stage (Figure 2F).

In summary, these results confirmed NSE4A to be a nuclear protein, as expected for a DNA repair factor, and revealed a dynamic expression pattern of NSE4A during sporogenesis, gametogenesis, embryogenesis, and endosperm development. The high levels of NSE4A during meiosis and in the proliferating fertilization products may be linked with its DNA repair function, for example, during meiotic crossing-over or to ensure genome integrity during the fast mitoses in embryo and endosperm.

### NSE4A Plays a Role in Seed Development

Prompted by NSE4 expression in seeds, we analyzed fertility of *nse4a* and *nse4b* mutants 2 weeks after pollination (Figures 3A and 3B). In contrast to the wild-type plants, siliques from *nse4a-1*/*NSE4A* heterozygotes produced 28.8% abnormal seeds (pale seeds representing delayed embryos and/or aborted seeds; *n* = 1402, Figures 3A and 3B). Fertility was even more impaired in

homozygous *nse4a-2* plants, with approximately one-half (53.4%) of the seeds developing normally, 22% showing early aborted ovules, and 24.6% showing abnormally large seeds with a glossy surface and liquid endosperm (*n* = 1008). Clearing of abnormal *nse4a-1* and *nse4a-2* seeds revealed that the embryos were arrested at the heart or heart-to-torpedo transition stages, respectively (Figure 3C; Supplemental Figure 6A). A *NSE4A* genomic construct could fully rescue the *nse4a-2* mutant seed phenotype (up to 96.5% normal seeds, *n* = 949), confirming that embryo unviability is a consequence of the loss of *NSE4A* function (Figures 3A and 3B). To test whether the increased frequency of abnormal seeds in *nse4a-1* heterozygous plants (28.8% observed versus expected 25%) is due to preferential transmission of the mutant allele or a partial gametophytic maternal effect, *nse4a-1*/*NSE4A* heterozygous plants were self-pollinated and reciprocally crossed to the wild-type plants. The frequency of late aborted seeds resulting from these crosses was scored (Supplemental Figure 6B). Reciprocal crosses resulted in 0.6 to 2.0% late aborted seeds, indistinguishable from the wild-type control, while self-pollinated *nse4a-1*/*NSE4A* heterozygous plants produced 23.9% late aborted seeds. These results indicate that *nse4a-1* is a zygotic embryo-lethal mutation. By contrast, and in agreement with the *NSE4B* expression pattern, *nse4b-1* and *nse4b-2* single mutants were fully fertile, while the *nse4a-2 nse4b-2* double mutant showed a similar phenotype as the *nse4a-2* single mutant (Figures 3A and 3B). Hence, *NSE4A* is required for normal seed development, while *NSE4B* is dispensable.

### NSE4A Is Involved in Somatic DNA Damage Repair

Next, we tested which of the Arabidopsis *NSE4* paralogs is involved in DNA damage repair. First, we scored for the
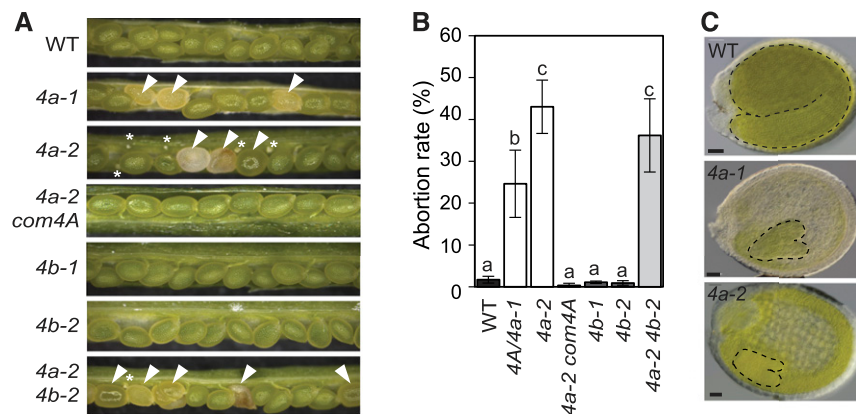


**Figure 3.** *NSE4A* Is Necessary for Seed Development.

**(A)** Seed phenotypes in the wild-type (WT), heterozygous self-pollinated *NSE4A*/*nse4a-1* (*4a-1*), homozygous *nse4a-2* (*4a-2*), homozygous *nse4a-2* complemented with genomic *NSE4A* locus (*4a-2 com4A*), *nse4b-1* (*4b-1*), *nse4b-2* (*4b-2*), and homozygous *4a-2 4b-2* double mutant plants. Abnormally developing seeds are indicated by white arrowheads. Nondeveloping ovules are indicated by white asterisks.
**(B)** Quantification of aborted seeds in the genotypes listed in **(A)**. Error bars indicate SD between means of three biological replicates. Each replicate was represented by one plant from which 140 to 300 seeds were analyzed. All plants were grown at the same time. Values marked with the same letter do not differ according to Duncan's multiple range test (P ≤ 0.05). WT, wild type.
**(C)** Equally old cleared wild-type (WT), pale self-pollinated *NSE4A*/*nse4a-1* (*4a-1*), and large *nse4a-2* (*4a-2*) seeds. Additional *nse4-2* seeds are shown in Supplemental Figure 6A. Embryos were outlined by black dashed lines for easier visibility. Bars = 50 μm.

transcriptional response of *NSE4A* and *NSE4B* to drug treatment using the promoter-GUS reporter lines (Figure 4). No induction was observed for *ProNSE4B:GUS* upon treatment with DNA damaging agents including zebularine (10 µM), which (similarly to the related drug 5-azacytidine; reviewed in Stingele and Jentsch, 2015; Tretyakova et al., 2015) generates enzymatic DNA–protein crosslinks by covalently trapping DNA Methyltransferase 1 class enzymes, and bleocin (25 nM), which causes DNA strand breaks (Figures 4A and 4C). By contrast, *ProNSE4A* became active throughout the entire meristematic zone and in the emerging lateral roots (Figures 4A and 4C), indicating that NSE4A is activated by different types of DNA damage. This transcriptional activation was accompanied by protein accumulation as indicated

**A**
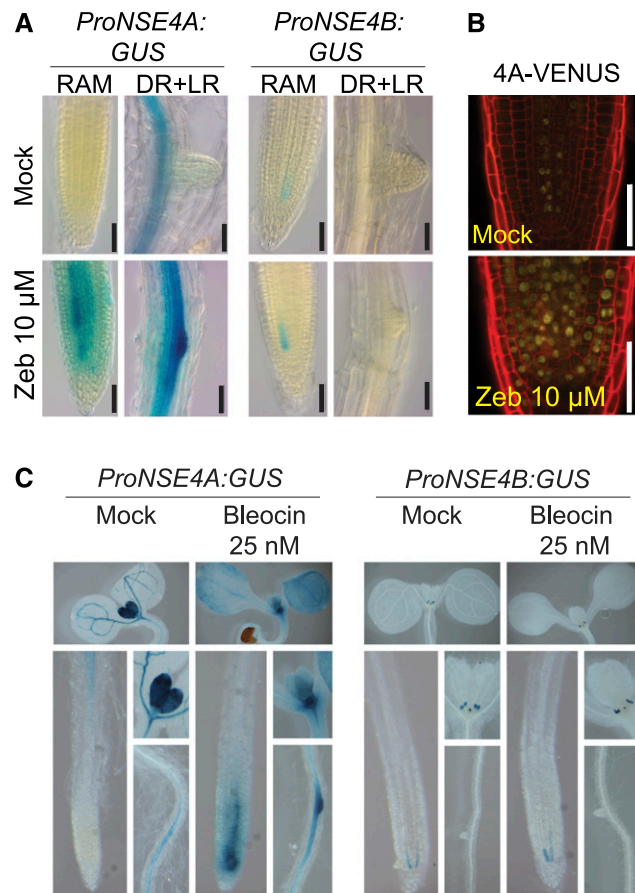


**Figure 4.** *NSE4A* Is Induced Upon DNA Damage Stimulus.

**(A)** Transcriptional response of the *ProNSE4A* and *ProNSE4B* promoters after 7 d of treatment with 10 µM zebularine (Zeb) in the root apical meristem (RAM) and differentiated root (DR) section with emerging lateral roots (LR). Scale bars = 50 µm.
**(B)** *nse4a-2 ProNSE4A:NSE4A:VENUS* (4A-VENUS) accumulation in the RAM under control conditions and with 10 µM zebularine (Zeb). Error bars = 50 µm.
**(C)** Transcriptional response of the *ProNSE4A* and *ProNSE4B* promoters to 25 nM bleocin treatment. Each composite image shows (from top to down and from left to right) the following: cotyledons and the first pair of true leaves, main root apical meristem, detail of the first pair of true leaves, and differentiated root zone.

by NSE4A-VENUS signals within a larger area of the root apical meristem of stressed reporter plants (Figure 4B).

Subsequently, we assessed the functional contribution of the *NSE4* genes to plant survival upon drug-induced DNA damage. To this aim, we monitored the growth of the wild-type, *nse4a-2* single mutant, *nse4a-2* complemented with *NSE4A* genomic construct (*ProNSE4A:NSE4A:TerNSE4A*), *nse4b* (both alleles), and *nse4a-2 nse4b-2* double mutant plants on media containing 10 µM zebularine, 50 nM bleocin, 10 µM MMC, or 1 mM hydroxyurea (HU; Figures 5A and 5B; Supplemental Figure 7). In a separate assay, we applied the DNA alkylating agent methyl methanesulfonate (MMS; Fig. 5C), which caused poor growth of the Arabidopsis *smc6b-3* (*mim-1*) mutant (Mengiste et al., 1999). As positive controls, we used the drug-sensitive *ATAXIA TELANGIECTASIA-MUTATED AND RAD3-RELATED* (*ATR*) signaling kinase mutant (*atr-2*), the *DNA LIGASE4* mutant (*lig4-2*), *WEE1 KINASE HOMOLOG* mutant (*wee1-1*), and mutants in the two SMC5/6 complex subunits, *SMC6B* (*smc6b-1*) and *HPY2* (*hpy2-2*; De Schutter et al., 2007; Ishida et al., 2009; Yuan et al., 2014; Liu et al., 2015). The *nse4b-1* and *nse4b-2* single mutants were not hypersensitive to any of the applied genotoxic treatments (Figures 5A to 5C). The *nse4a-2* single and *nse4a-2 nse4b-2* double mutants were indistinguishable from the wild type under MMC, bleocin, and HU stress, but they were strongly hypersensitive to zebularine and MMS (Figures 5A to 5C). By contrast, *smc6b-1* was also hypersensitive to MMC treatment, which could be due to the fact that *nse4a-2* is only a partial loss-of-function allele. To test for effect on homologous recombination (HR) rates, we generated *nse4a-2 nse4b-2* double mutants carrying the reporter N1DC1 No. 11 (B11) with 566 bp overlap of GUS recombination substrate in direct orientation (Puchta et al., 1995). The plants were grown for 10 d in media containing low amounts of zebularine (1.25 and 2.5 µM) to avoid lethality. We used multiple independent lines of each analyzed genotype, which showed a zebularine dose-dependent increase in HR rate, but no significant differences between the wild-type, *nse4a-2*, and *nse4b-2* lines (Figure 5D). This result differs from those published for *hpy2* and *smc6* mutants, which showed reduced HR rates (Mengiste et al., 1999; Watanabe et al., 2009; Yuan et al., 2014). On the one hand, this may suggest that NSE4 proteins are not controlling single strand annealing type of HR in Arabidopsis. On the other hand, these results should be interpreted with caution because *nse4a-2* is not a null allele and *nse4b* mutants are not sensitive to DNA damage treatments.

Inhibition of root growth in response to DNA damage is frequently accompanied by increased cell death. Therefore, we monitored the amount of dead cells using the propidium iodide (PI) uptake assay in control and 20 µM zebularine-treated plants (Figure 5E). While there were no or few dead cells in the wild-type and *nse4b-2* plants, *nse4a-2* single and *nse4a-2 nse4b-2* double mutant plants showed a drastic increase upon zebularine treatment. The drug sensitivity phenotype (growth and cell death) of *nse4a-2* to zebularine is directly due to the loss of *NSE4A* activity as shown by complementation using an *NSE4A* genomic construct (Figures 5A, 5B, and 5E). We noticed that the root meristem was partially disorganized in zebularine-treated *nse4a-2* plants. Therefore, we estimated the meristem size by counting the number of cells in the cortex layer between the quiescent center
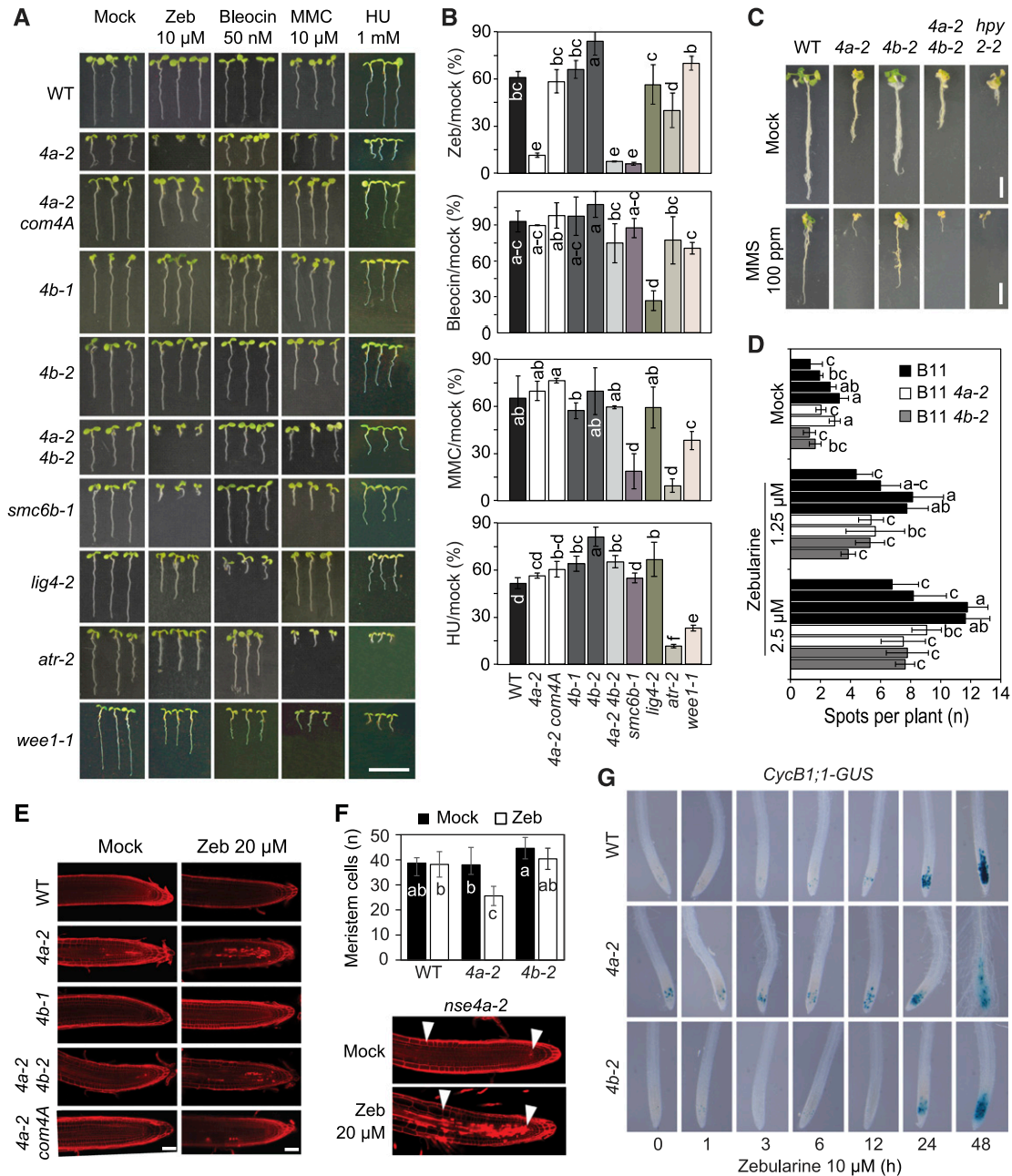
**Figure 5.** *NSE4A* Is Involved in Somatic DNA Damage Repair.

**(A)** Sensitivity to genotoxic stress. The wild-type (WT), *nse4a-2* (*4a-2*), *nse4b-1* (*4b-1*), *nse4b-2* (*4b-2*), *nse4a-2 nse4b-2* (*4a-2 4b-2*), *nse4a-2* complemented with genomic *NSE4A* locus (*4a-2 com4A*), *smc6b-1*, *lig4-2*, *atr-2*, and *wee1-1* plants were germinated and maintained for 1 week on 10 μM zebularine (Zeb), 50 nM bleocin, 10 μM MMC, or 1 mM HU. Bar = 10 mm.

**(B)** Quantitative data for **(A)** calculated as the relative root length under drug versus control conditions. Error bars represent sᴅ between means of three biological replicates. The replicates were grown on separate screening plates, and each contained at least 25 plants. Values marked with the same letter do not differ according to Duncan's multiple range test (P ≤ 0.05). WT, wild type.

**(C)** Sensitivity to MMS. Representative phenotypes of the wild-type (WT), *4a-2*, *4b-2*, *4a-2 4b-2* double mutant, and *hpy2-2* plants grown for 1 week in control liquid medium and then for 3 weeks in control and 100 ppm MMS-containing media. Bar = 10 mm.

**(D)** Analysis of DNA damage repair by homologous recombination using B11 reporter line in the wild-type (WT), *4a-2*, and *4b-2* backgrounds. Identically colored columns represent individual lines obtained from segregating hybrid populations. Error bars represent mean of three biological replicates, each with at least 30 plants. Values marked with the same letter do not differ according to Duncan's multiple range test (P ≤ 0.05).

**(E)** Cell death assay. PI-stained roots from living Arabidopsis seedlings treated without (Mock) and with 20 μM zebularine (Zeb) for 24 h. WT, wild type.

and the differentiation zone (Figure 5F). The wild-type and *nse4b-2* roots contained 38 to 45 cells, and this number did not change significantly after 24 h of 20 μM zebularine treatment (analysis of variance, post hoc Duncan's test, P > 0.05). By contrast, *nse4a-2* showed a significant 31% reduction to 26 cells upon zebularine treatment. To test the effect of the mutation on cell cycle regulation, we introduced a G2/Mitosis DNA damage reporter, which utilizes a translational fusion between CyclinB1;1 and GUS (Colón-Carmona et al., 1999), into *nse4a-2* and *nse4b-2* mutant backgrounds. The chimeric protein accumulates specifically in the G2 phase of cycling cells and is destroyed at the onset of mitosis, resulting in a loss of the signal. Double homozygous lines were exposed to 10 μM zebularine for up to 48 h, and the domain of GUS expression was monitored (Figure 5G). The *nse4a-2* roots showed an increased number of GUS-positive cells already at 0 h, indicating a prolonged G2 phase. After 48 h of treatment, meristems of *nse4a-2* plants were damaged, as indicated by an abnormal root morphology and root hairs emerging close to the root tips. The response in *nse4b-2* and the wild type was slower, less severe, and similar between the two (Figure 5G).

Collectively, these results demonstrate that *NSE4A* responds to genotoxic stress, is likely involved in DNA repair of zebularine-induced DNA–protein crosslinks, and is required to promote cell division in response to this genotoxic drug, possibly to actively propagate cells after repair.

### Loss of *NSE4A* Function Causes Upregulation of DNA Damage Repair and Immune Response Genes

We analyzed the effect of the *nse4a-2* mutation on gene expression by RNA sequencing using dissected shoot apices from the 10-d-old wild-type and *nse4a-2* plants treated without (mock) and with 20 μM zebularine for 24 h (Figure 6; Supplemental Data Set 4). In mock-treated *nse4a-2*, we identified 555 significantly upregulated genes and 181 significantly downregulated genes relative to the mock-treated wild type (Figure 6A; DESeq, adjusted P < 0.05; the same parameters apply to the whole section). In zebularine-treated wild-type plants, we found 446 significantly upregulated genes and 183 significantly downregulated genes, that is, many more than we identified in a previous study (Liu et al., 2015). This difference is most likely due to the treatment in liquid media, allowing for a more intense uptake of zebularine compared with the previously used solid media. Zebularine treatment of *nse4a-2* plants had the strongest effect, leading to upregulation of 1374 genes and downregulation of 773 genes compared with mock-treated *nse4a-2* control plants. Upregulated genes included several prominent DNA damage repair markers (Figure 6B). These data suggest that the SMC5-SMC6 complex is not required

for transcriptional upregulation of DNA damage repair genes, but loss of its functionality triggers a more intense DNA damage response (Figure 6B).

Previous microarray-based expression analysis of *sni1-1* suggested a link between function of the SMC5/6 complex and immune responses (Mosher et al., 2006). Comparison of the transcriptomes from *nse4a-2* and *sni1-1* mutants revealed 82 (5.8%) commonly upregulated and 6 (0.5%) commonly downregulated genes (Figure 6C; Supplemental Data Set 5). The upregulated genes were mainly associated with stress responses, defense responses to (biotic) stimuli, and responses to other organisms (Figure 6D; Supplemental Table 3), which was described for *SNI1* (Mosher et al., 2006; Yan et al., 2013) but is new information for *NSE4A*. The upregulated genes in *nse4a-2* plants included *PATHOGENESIS-RELATED GENE2* (*PR2*; also known as *BETA-1,3-GLUCANASE2*), *PR4*, *PR5*, and several *TOLL/ INTERLEUKIN-1 RECEPTOR-NUCLEOTICDE BINDING SIGNAL-LEUCINE RICH REPEAT* genes (At5g46490, *WHITE RUST RESISTANCE4*, At3g44630; Figure 6E; Supplemental Data Set 5). This indicates that mutations affecting the SMC5/6 complex cause constitutive expression of immune response genes and lead to activation of other DNA damage repair pathways, most likely due to accumulation of spontaneous DNA damage.

### NSE4A and NSE4B Interact with the Same SMC5/6 Complex Subunits

In plants, the architecture of SMC5/6 complex remains unknown. Based on fungal and animal models, we assume that NSE4 may act as a central subunit interacting with SMC5 and SMC6, and possibly several other NSEs (Duan et al., 2009; Hudson et al., 2011). To test whether this hypothesis holds true for both NSE4 paralogs, we performed yeast two-hybrid (Y2H) assays. The assay conditions were optimized using the positive (T+53) and the negative (T+lam C) controls, and we suppressed protein autoactivation by adjusting the 3-amino-1,2,4-triazole (3-AT) concentrations (Figure 7A; Supplemental Figure 8; Supplemental Table 4). As a control, we confirmed the interaction of SMC6A and SMC6B hinges with the SMC5 hinge (Figure 7A). Subsequently, we tested for interactions of full-length SMC5 or SMC6 with NSE4A and NSE4B. While the interaction between both NSE4 paralogs and SMC5 was positive (Figure 7A), we did not observe yeast growth when testing interactions with SMC6A and SMC6B. This remained true even after switching the tag positions (N- and C-terminal positions) and extensive optimization (Supplemental Figure 8). Within the NSE1-NSE3-NSE4 subcomplex, we measured positive interactions of both NSE4 paralogs with NSE3 and confirmed (Li et al., 2017) the interaction of NSE1 with NSE3 (Figure 7A). However, we did not detect interactions between

---

**Figure 5.** (continued).

**(F)** Meristem size estimation. Plants from **(E)** were used to estimate the number of cells within the root apical meristem (indicated by white arrowheads). Error bars in graph indicate SD among primary roots from 5 to 12 analyzed plants per each genotype. All plants were grown at the same time. Values marked with the same letter do not differ according to Duncan's multiple range test (P ≤ 0.05). WT, wild type; Zeb, zebularine.

**(G)** G2/M cell cycle progression in *nse4a-2* and *nse4b-2* analyzed by *ProCycB1;1:CycB1;1:GUS* (*CycB1;1-GUS*) after exposure to 10 μM zebularine for the indicated number of hours.
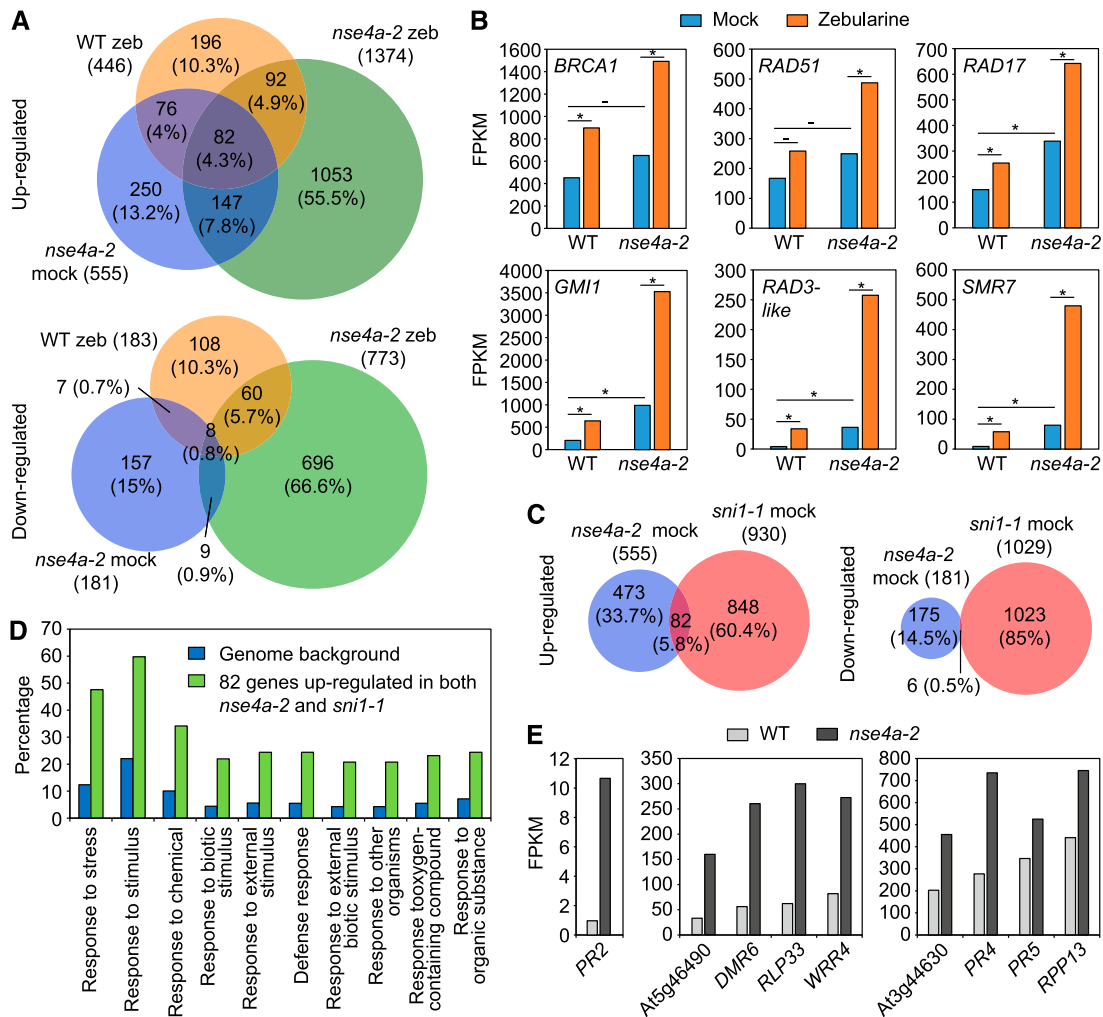
**Figure 6.** Transcriptome Analysis of *nse4a-2* Plants.

**(A)** Venn diagrams of genes significantly (DESeq, adjusted P < 0.05) up- and downregulated in dissected shoot apices of the 20 µM zebularine (zeb)–treated wild-type (WT zeb/WT mock), mock-treated *nse4a-2* (*nse4a-2* mock/WT mock), and 20 µM zeb-treated *nse4a-2* (*nse4a-2* zeb/*nse4a-2* mock) plants. The data are based on two RNA sequencing replicates.

**(B)** mRNA abundance of DNA damage repair marker genes expressed as fragments per kilobase per million of reads (FPKM) based on data shown in **(A)**. Asterisks and dashes indicate statistically significant and nonsignificant, respectively, differences between groups indicated by horizontal bar in DESeq (adjusted P-value < 0.05). WT, wild type; *BRCA1*, *BREAST CANCER SUSCEPTIBILITY1*; *RAD51*, *RADIATION SENSITIVE51*; *RAD17*, *RADIATION SENSITIVE17*; *GMI1*, *GAMMA-IRRADIATION AND MITOMYCIN C INDUCED1*; RAD3-like,RADIATION SENSITIVE3-like, At1g20750; *SMR7*, *SIAMESE-RELATED7*.

**(C)** Venn diagrams of significantly up- and downregulated genes in *nse4a-2* (see **[A]**) and *sni1-1* (*sni1-1* mock/wild type (WT) mock; ATH1 expression microarrays, adjusted P < 0.05) plants.

**(D)** Gene ontology (GO) term analysis of 82 genes significantly upregulated in both *nse4a-2* and *sni1-1* (see **[C]**) using agriGO v2.0. Top 10 GO term categories are shown as input relative to Arabidopsis genomic background/reference. The full list of significant GO terms is available in Supplemental Table 3.

**(E)** Examples of significantly (DESeq, adjusted P-value < 0.05) upregulated defense-related genes in dissected shoot apices of mock-treated *nse4a-2* plants. *DMR6*, *DOWNY MILDEW RESISTANT6*; *RLP33*, *RECEPTOR LIKE PROTEIN33*; *WRR4*, *WHITE RUST RESISTANCE4*; *RPP13*, *RECOGNITION OF PERONOSPORA PARASITICA 13*.

NSE4A or NSE4B and NSE1. To validate the interactions identified by Y2H, we performed bimolecular fluorescence complementation (BiFC) assays in *Nicotiana benthamiana* and analyzed signals using confocal microscopy (Figure 7B). In all cases, the signals were localized to the nucleus and confirmed that both NSE4A and

NSE4B are able to interact with SMC5 and NSE3. Moreover, we tested protein–protein interactions using coimmunoprecipitation (co-IP) assays in *N. benthamiana* and validated (1) the interactions of the SMC5 hinge with the hinges of SMC6A and SMC6B, (2) the interaction of NSE3 with NSE4A and NSE4B, and (3) the
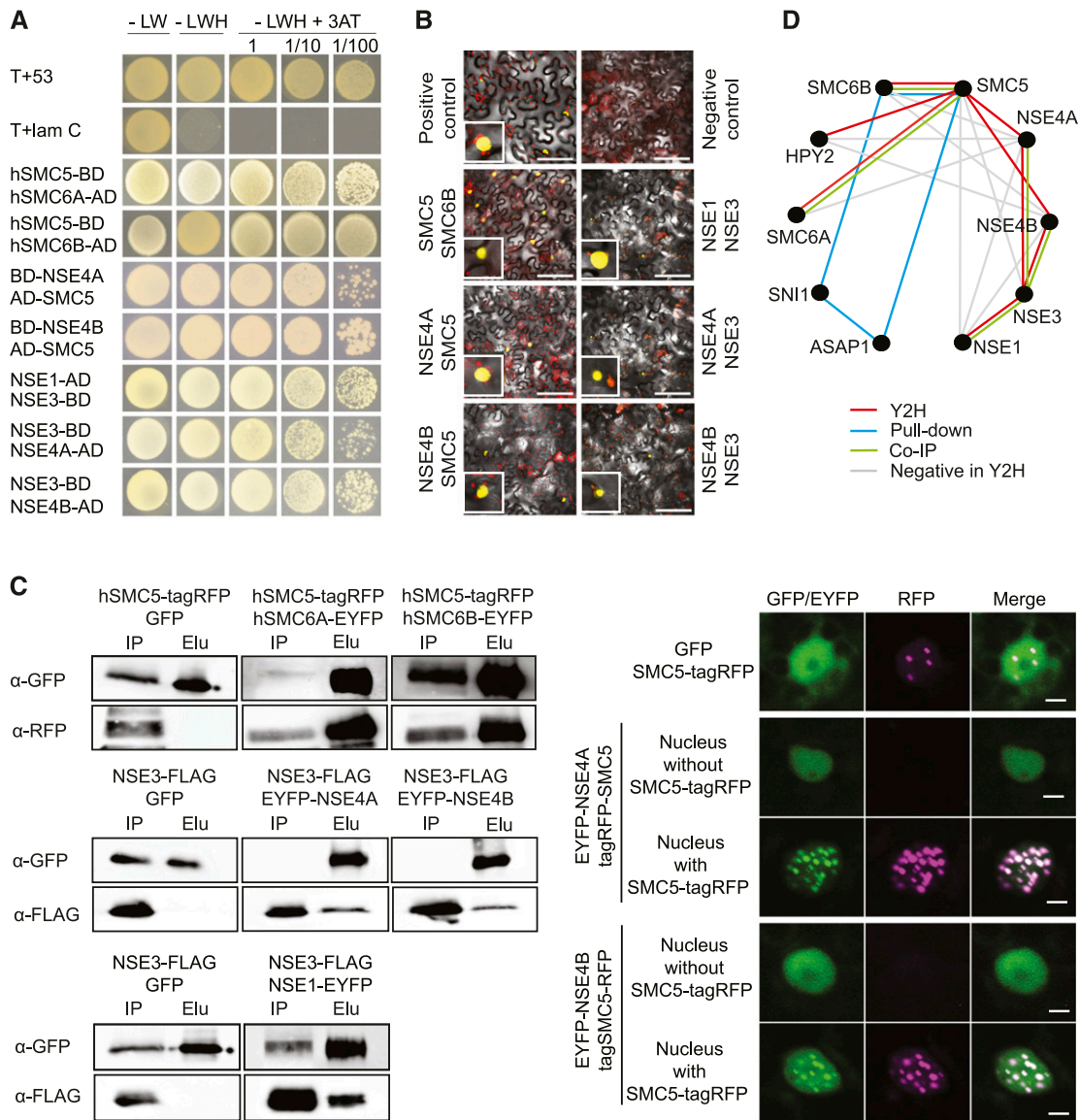
**Figure 7.** Analysis of Protein–Protein Interactions.

**(A)** Y2H assays. T+53, positive control and T+lam C, negative control. Domain position before/after the gene name indicates N- or C-terminal fusions, respectively. Autoactivation controls, negatively tested combinations, and used 3-AT concentrations are provided in Supplemental Figure 8 and Supplemental Table 4. -LW, without leucine and tryptophan; -LWH, without leucine, tryptophan and histidine; h, hinge domain, BD, binding domain, AD, activation domain.

**(B)** BiFC validation of interactions indicated by Y2H. Insets show nuclei with positive signals. Bars = 50 μm.

**(C)** co-IP and colocalization assays. Right panel displays co-IP analysis. Whole blots are shown in Supplemental Figure 9. Right panel shows changes in EYFP-NSE4A and EYFP-NSE4B localization after addition of SMC5-tagRFP. Elu, elution (proteins collected by green fluorescent protein trapping); GFP, GREEN FLUORESCENT PROTEIN trapping; RFP, RED FLUORESCENT PROTEIN; IP, input (total protein extract); h, hinge domain.

**(D)** Model of protein–protein interactions within Arabidopsis SMC5/6 complex based on Y2H and BiFC (red lines), pull-down (Yan et al., 2013), and co-IP (green lines) experiments. Negatively tested combinations in Y2H are indicated by gray lines. Interaction between HPY2 and SMC5 was published previously (Xu et al., 2013).

interaction of NSE3 with NSE1 (Figure 7C, left; Supplemental Figure 9). We could not evaluate the interactions of NSE4A and NSE4B with the full-length SMC5 protein using co-IPs because, despite extensive optimization, SMC5 did not reach detectable levels following transfection in *N. benthamiana* leaves as assayed by protein gel blotting. However, the presence of tagRED FLUORESCENT PROTEIN (tagRFP)-SMC5 modified the nuclear distribution of both NSE4A-ENHANCED YELLOW FLUORESCENT PROTEIN (EYFP) and NSE4B-EYFP from a dispersed to a speckled pattern (Figure 7C, right).

In summary, the results from Y2H, BiFC, and co-IP assays together with published data allow us to conclude that individual Arabidopsis SMC5/6 complex subunits interact and that SMC5 recruits NSE4A and NSE4B into speckled domains in the nucleus (Figure 7C). Based on these experiments, we developed a model for interactions between SMC5/6 complex subunits in Arabidopsis (Figure 7D).

### The NSE4B Protein Can Partially Substitute NSE4A Protein Functions

The *NSE4A* and *NSE4B* paralogs show little overlap in their expression patterns and loss-of-function phenotypes. To test whether *NSE4A* and *NSE4B* also diverged functionally, we developed a promoter swap construct consisting of the *NSE4B* genomic coding sequence (CDS) under the control of the *NSE4A* promoter (*ProNSE4A*:*GenomicNSE4B*:*TerNSE4B*). This construct was transformed into homozygous *nse4a-2* plants, and individuals heterozygous or homozygous for the promoter swap construct were selected in the T2 generation and tested for zebularine sensitivity in the T3 generation. While the control *nse4a-2* plants were strongly hypersensitive, several independent promoter swap lines showed rescue, albeit incomplete, of the drug sensitivity phenotype, with average roots length being intermediate between those of *nse4a-2* and the wild-type plants (Figures 8A and 8B).

In addition, the broader expression domain of *NSE4B* in the promoter swap lines was able to rescue the seed abortion phenotype of *nse4a-2* (Figures 8C and 8D). Furthermore, *NSE4B* expression in the *nse4a-1* background allowed the recovery of homozygous *nse4a-1* plants (24% viable *nse4a-1*/*nse4a-1* plants in the progeny of a *NSE4A*/*nse4a-1*;*ProNSE4A*:*GenomicNSE4B*:*TerNSE4B* segregating parent; *n* = 92, Supplemental Tables 5 and 6).

Taken together, these results demonstrate that NSE4A and NSE4B have similar biochemical activities that are fully exchangeable during seed development but only partially in DNA damage responses.

### DISCUSSION

The SMC5/6 complex plays a crucial role in the maintenance of genome stability in eukaryotes (De Piccoli et al., 2009; Kegel and Sjögren, 2010; Jeppsson et al., 2014b; Diaz and Pecinka, 2018). Some of its subunits remain poorly characterized in plants, including the two *NSE4* homologs. Here, we demonstrate that *NSE4A* is involved in preserving genome stability and controls seed development. *NSE4B* is barely active during normal development and nonresponsive to drug-induced genotoxic stress.

### NSE4A Is an Essential Gene in Arabidopsis

The *NSE4* paralogs of Arabidopsis originate from the whole-genome duplication event (α) that occurred ~47 MYA in Brassicaceae (Kagale et al., 2014). Surprisingly, there were at least two *NSE4A* copies in all vascular plants analyzed, with the highest number of six copies in *Oryza sativa*. The *NSE4* amplifications are family
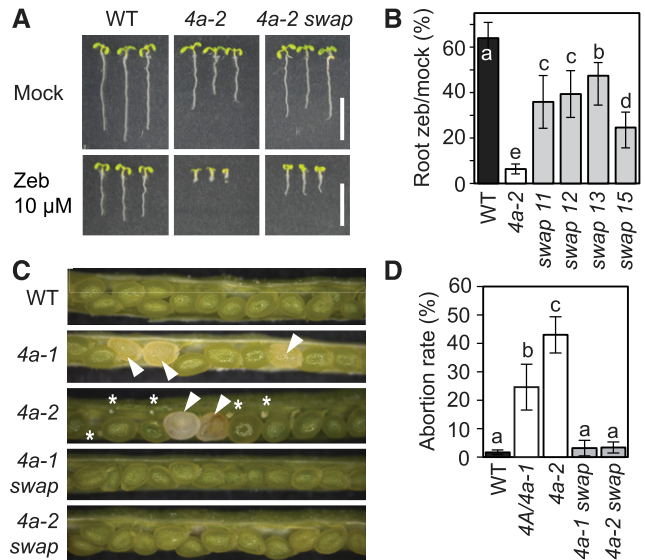


**Figure 8.** Analysis of *NSE4B* Functions.

**(A)** Zebularine (Zeb) hypersensitivity assay. Wild-type (WT), *nse4a-2* (*4a-2*), and *nse4a-2* complemented with *ProNSE4A*:*GenNSE4B*:*TerNSE4B* (*4a-2 swap*) line 13 were germinated and kept on control and 10 μM Zeb-containing media for 1 week. Bar = 10 mm.

**(B)** Quantitative data for root length of zebularine (zeb)–treated versus control plants as described in **(A)**. Lines 11, 12, 13, and 15 represent independent promoter swap transgenic lines. Error bars indicate sᴅ between the means from two biological replicates. Each replicate consisted from at least 20 plants per line grown on separate screening plates at different times. Values marked with the same letter do not differ according to Duncan's multiple range test (P ≤ 0.05). WT, wild type.

**(C)** Analysis of seed development phenotypes in the wild type (WT), heterozygous *NSE4A*/*nse4a-1* (*4a-1*), and *4a-2*. The two bottom pictures show homozygous *nse4a-1* and *nse4a-2* containing homozygous promoter swap line 13 (*4a-1 swap* and *4a-2 swap*). White arrowheads indicate aberrantly developing seeds and asterisks aborted ovules.

**(D)** Quantification of abortion rates in the genotypes described in **(C)**. Error bars indicate sᴅ between means of three biological replicates (plants), each with at least 300 scored seeds. Values marked with the same letter do not differ according to Duncan's multiple range test (P ≤ 0.05). WT, wild type.

specific and much more frequent than duplications of any other SMC5/6 complex members in plant genomes (reviewed in Diaz and Pecinka, 2018). Our data from Arabidopsis and published data from humans (Hudson et al., 2011) suggest that at least some of these duplicated copies differ in their expression domains. We found that both NSE4A and NSE4B can interact with the core subunits SMC5 and NSE3, but not with NSE1, with the latter two representing members of the NSE1-NSE3-NSE4 subcomplex (Palecek and Gruber, 2015). However, in spite of extensive optimization, we did not detect interactions of the NSE4 proteins with SMC6B. This interaction is very likely to exist in Arabidopsis but seems particularly difficult to confirm as indicated by previous studies in *Saccharomyces cerevisiae* and *S. pombe* (Palecek et al., 2006; Duan et al., 2009; J. Palecek, personal communication). This is possibly caused by a steric hindrance due to the specific conformation of SMC6 and NSE4 proteins or the absence of an activating and/or stabilizing component.

A strong *nse4a* mutation was homozygous lethal, and self-pollinated heterozygotes showed 28.8% seed abortion. This resembles the phenotypes of *smc5*, *nse1*, *nse3*, and *asap1* mutants and the *sm6a smc6b* double mutant, which show embryonic or cotyledon-stage seedling death in Arabidopsis (Watanabe et al., 2009; Xu et al., 2013; Yan et al., 2013; Li et al., 2017). However, we also found a hypomorphic *nse4a-2* allele, which likely produces a protein with a modified C terminus. This allele alleviates the problem of homozygous lethality encountered in the loss-of-function allele *nse4a-1*, thereby enabling the analysis of *NSE4A* functions during plant development and genotoxic stress. Its phenotypes partially resemble those of *HPY2* and *SNI1* mutants, which survive but are strongly affected in development and fertility (Li et al., 1999; Huang et al., 2009; Ishida et al., 2009).

### *NSE4A* Is Involved in Sporogenesis, Gametogenesis, and Seed Development

We observed prominent and dynamic expression of *NSE4A* during Arabidopsis reproductive development. In the male gametophyte, *NSE4A* was expressed in the generative cell lineage but absent in the vegetative cell. This is consistent with the observation that the sperm nucleus is rich in the components of active chromatin control, while the vegetative nucleus has lost multiple repressive chromatin modifications and will no longer divide (Schoft et al., 2009; Slotkin et al., 2009; Abdelsamad and Pecinka, 2014). However, the function of *NSE4A* in pollen development remains unknown. Possibly, NSE4A secures a faster or more accurate response, which is not detected under laboratory conditions, upon environmental challenges affecting genome integrity in the germline.

*NSE4A* is also broadly expressed in ovule primordia, with a notable accumulation in the female meiocyte. Thus, besides its role in male meiosis (Liu et al., 2014), the SMC5/6 complex may play a role during female meiosis, possibly in the process of DNA replication, meiotic recombination, or DNA damage repair. During embryo sac development and early seed development, *NSE4A* was expressed in synergids and the central cell and later in the embryo and the syncytial and chalazal endosperm. *NSE4A* expression at these stages may be interpreted as a functional requirement for genome integrity safeguarding processes, which involve DNA repair as a consequence of the challenges posed by rapid DNA replication and chromatin dynamics in these tissues (Baroux et al., 2007; Baroux and Autran, 2015). Genome integrity is necessary to ensure the proper differentiation and functioning of the progeny and to avoid the propagation of genetic mutations. In addition, but not exclusively, the high levels of *NSE4A* in the syncytial endosperm may play a role in the detoxification of endogenously occurring replication-derived toxic DNA structures. DNA replication produces a high frequency of inter-twining between nascent chromatids, DNA supercoils, and X-shaped toxic DNA replication intermediates, which all require (to different extents) SMC5/6 functions for resolution (Jeppsson et al., 2014a; Menolfi et al., 2015; reviewed in Diaz and Pecinka, 2018).

While SMC5/6 complex null mutations lead to early seed abortion, the hypomorphic *nse4a-2* mutant produced large glossy seeds with liquid endosperm, which turned brown at later stages and aborted. Seed phenotypes similar to *nse4a-1* or *nse4a-2* were reported for *nse1*, *nse3*, and *mms21/hpy2* mutants (Liu et al., 2014; Li et al., 2017). Studies in *S. cerevisiae* revealed that the SMC5/6 complex is loaded by the Sister chromatid cohesion protein 1 subunit of the cohesin complex to specific sites during DNA replication (Jeppsson et al., 2014a). This could explain the similarity of SMC5/6 complex and cohesin mutant seed phenotypes and indicates that both complexes cooperate during seed development. This may be supported by the identification of cohesin, and also condensin, mutants in a screen focusing on aberrant seed development (Liu et al., 2002; Tzafrir et al., 2002) and underlines the importance of maintaining genome stability during seed development (reviewed in Diaz and Pecinka, 2017).

### *NSE4A*, but Not *NSE4B*, Is Required for Resistance to Genotoxic Stress

The functions of the SMC5/6 complex are widely associated with the maintenance of genome stability (Kegel and Sjögren, 2010; Wu and Yu, 2012; Jeppsson et al., 2014b); however, it was not clear which of the Arabidopsis *NSE4* paralogs confers this function. We observed activation of *NSE4A*, but not *NSE4B*, in response to genotoxic treatments with drugs inducing various types of DNA damage. In addition, the viable and phenotypically almost wild-type *nse4a-2* plants were hypersensitive to the cytidine analog zebularine and the alkylating agent MMS, but not to other treatments. Lack of sensitivity to bleocin, MMC, and HU could be caused by the fact that the mutation we analyzed is not a complete loss-of-function allele and/or that such damages can be processed by SMC5/6-independent pathways. We have previously shown that *smc6b* mutants are hypersensitive to zebularine-induced damage (Liu et al., 2015). This suggests that the SMC5/6 complex is essential for detoxification from complex toxic structures, such as zebularine-induced DNA damage. DNA repair in response to zebularine treatment is mediated both by ATAXIA TELANGIECTASIA-MUTATED and ATR kinases (Liu et al., 2015), which are known to phosphorylate proteins at Ser followed by Gln or Thr followed by Gln motifs (Awasthi et al., 2015). NSE4A contains two adjacent Thr-Gln motifs at amino acids 361 to 365 (TQDTQ), which makes it a good candidate for a direct target of phosphorylation by ATM and/or ATR.

Recent studies from nonplant models suggest that the SMC5/6 complex acts as an ATP-dependent intermolecular linker, which helps resolving toxic DNA structures at late-replicating sites and also prevents recombination between nonhomologous sequences (Chiolo et al., 2011; Kanno et al., 2015; Menolfi et al., 2015). In Arabidopsis, the SMC5/6 complex promotes the association of sister chromatids and is required for normal levels of homologous recombination (Mengiste et al., 1999; Hanin et al., 2000; Watanabe et al., 2009; Yuan et al., 2014). In addition to its role in somatic DNA damage repair, there is emerging evidence that the SMC5-SMC6 complex also plays a role in immune responses (Yan et al., 2013) and meiosis (Yuan et al., 2014). Our data indirectly support a meiotic role of NSE4A as it strongly accumulates in female meiocytes. However, the exact molecular

mechanism of genome maintenance by the SMC5/6 complex remains unknown.

## NSE4B and NSE4A Have Primarily Diversified Transcriptionally, and NSE4B Is Not Responsive to DNA Damage

In Arabidopsis, the functions of *NSE4B* are less clear than those of *NSE4A*. *NSE4B* single mutants are morphologically indistinguishable from the wild type and do not worsen the phenotype of a weak *nse4a* mutant. We found that *NSE4B* is silenced throughout most of development, except for a small domain in the root apical meristem, leaf stipules, and the embryo up to the globular stage. Based on the results of in silico analyses, which revealed an extensive coverage of the *NSE4B* locus by histone H3 Lys-27 trimethylation, we hypothesize that *NSE4B* is controlled by the *Polycomb* Repressive Complex 2 (reviewed in Mozgova and Hennig, 2015). To explore *NSE4B*'s function in the nonsilenced state, we swapped its promoter with that of *NSE4A* and tested whether *NSE4B* expressed in the pattern of *NSE4A* can complement the *nse4a* phenotypes. The seed abortion phenotype was fully complemented, but we found only a partial rescue under DNA damaging conditions. This points to the dual function of the SMC5/6 complex described in budding yeast (Menolfi et al., 2015): a DNA damage-independent function during DNA replication and a DNA damage-dependent function in DNA repair. Both NSE4A and NSE4B seem capable of performing the first function, while DNA damage repair can be done only by NSE4A in Arabidopsis.

## METHODS

### Plant Material

The Arabidopsis (*Arabidopsis thaliana*) wild type and mutants were in the Col background: *nse4a-1* (SALK_057130), *nse4a-2* (GK-768H08), *nse4b-1* (SAIL_296_F02), *nse4b-2* (GK-175D10), *smc6b-1* (SALK_ SALK_101968C), *hpy2-2* (SAIL_77_G06), *atr-2* (SALK_032841C), *wee1-1* (GK-270E05), and *lig4-2* (SALK_044027C). We also used a cyclin-GUS line containing the *ProCYCB1;1:CYCB1;1:GUS* construct (Colón-Carmona et al., 1999) and the B11 line containing an intramolecular type of HR substrate (Puchta et al., 1995). For promoter reporter constructs, regions 18,943,545 to 18,941,640 and 7,260,588 to 7,258,919 bp upstream of the *NSE4A* and *NSE4B* transcription start sites, respectively, were PCR amplified, cloned into *pDONOR221*, and recombined into the binary Gateway vector pGWB553 containing the *uidA* gene encoding GUS. The final plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 and then into Arabidopsis Col using the floral dip method (Zhang et al., 2006). T1 generation seeds were screened on one half Murashige and Skoog (MS) plates containing 25 μg/L hygromycin B (Duchefa Biochemie), and resistant plants were transferred to soil. T2 populations with ~75% resistant seedlings, indicating single locus T-DNA insertions, were considered for further analyses. For promoter swap experiments, the *NSE4A* promoter and genomic region of *NSE4B* were PCR amplified and cloned into the *pGWB550* vector by MultiSite Cloning Gateway (Thermo Fisher Scientific). The construct was transformed into the *nse4a-2* background using the floral dip method. To construct the NSE4A-fluorescent protein translational fusion, the *NSE4A* promoter, CDS, terminator, *VENUS* N-terminal tag, and a BASTA resistance cassette were cloned using

Gibson assembly (New England Biolabs) into *pGGA000*, *pGGC000*, *pGGE000*, *pGGB000*, and *pGGF000*, respectively, to generate entry clones. The Greengate cloning reaction was performed as described previously (Lampropoulos et al., 2013), and the multi entry cassette was assembled into the *pAGM4723* backbone. *nse4a-2* mutant plants were transformed with this construct using the floral dip method. For *nse4a-2* complementation analysis, the *NSE4A* promoter and genomic region of *NSE4A* were PCR amplified and cloned into the *pGWB550* vector by MultiSite Cloning Gateway (Thermo Fisher Scientific). Plant transformation and screening of transformants were performed exactly as for the promoter swap experiment. Plants were emasculated ~48 h prior to pollination in crossing experiments.

### Phylogenetic Analysis and Shadowing

NSE4 protein sequences were retrieved from the National Center for Biotechnology Information and Phytozome (Supplemental Table 1). The protein alignment was performed using the MUSCLE algorithm (Edgar, 2004), and the resulting alignment was submitted to Gblocks (Castresana, 2000). Curation and selection of aligned blocks were performed in Gblocks using less stringent parameters. Bootstrap probabilities for each node were calculated with 100 replicates. Original sequences, alignments, and blocks are provided as Supplemental Data Sets 1, 2, and 3, respectively.

Promoter sequences from all analyzed species were retrieved from Phytozome (Supplemental Table 2). Promoter regions of *NSE4A* and *NSE4B* were submitted individually to mVISTA (Frazer et al., 2004), and sequence conservation was calculated using LAGAN program (Brudno et al., 2003). The Arabidopsis sequences were used as references for pairwise comparisons (Supplemental Figure 2).

### Plant Growth Conditions and Drug Treatments

For genotyping, crossing, and seed production plants were grown in 7 × 7-cm pots filled with peat bog in a climatic chamber under controlled long-day conditions (at 16 h with an ~200 μmol m$^{-2}$ s$^{-1}$ light intensity and 21°C during day; 8 h at 19°C during night) with standard 70% humidity.

For in vitro experiments, sterilized seeds were evenly spread on sterile one half Murashige and Skoog (MS) medium with or without zebularine (Sigma-Aldrich), MMC (Duchefa Biochemie), bleocin (Calbiochem), and HU (Sigma-Aldrich) in concentrations specified in the text and grown at 16 h with 150 μmol m$^{-2}$ s$^{-1}$ light:8 h dark at 21°C. Seven-day-old plants were used for root length measurements. For MMS experiment, sterilized seeds were grown in one half MS medium for 5 d and then transferred to liquid one half MS medium with and without 100 ppm MMS, and grown for 26 d. Roots from 20 to 25 seedlings per genotype were straightened, and in total three replicates were performed. For RNA sequencing, seeds were germinated on drug-free on half MS solid medium, and 9-d-old plants were carefully transferred to liquid one half MS medium with or without 20 μM zebularine. After 24 h, plants were washed with drug-free liquid one half MS medium; their leaves, hypocotyl, and roots were removed; and shoot apices were flash-frozen in liquid nitrogen for later use.

### Nucleic Acid Isolation, cDNA Synthesis, and PCR

For DNA isolation, leaf material of plants at the rosette stage was harvested, and DNA was isolated using the DNeasy Plant Mini kit (QIAGEN), following the manufacturer's instructions. For RNA isolation, floral buds were collected, shock-frozen in liquid nitrogen, and kept at −80°C until use. Total RNA isolation was performed with QIAzol (QIAGEN), and the RNA integrity was assessed by formaldehyde agarose gel electrophoresis. cDNA synthesis was performed from 1 μg of total RNA as starting material, using the RevertAid First Strand cDNA Synthesis kit (Thermo Fisher Scientific) with

oligo(dT) primers according to manufacturer's instructions. Primers used in this study are provided in Supplemental Tables 7 and 8. For 3′ RACE PCR, we performed in total four nested PCR reactions using the primer combinations listed in Supplemental Table 8. The first PCR was performed using a 1/100 (v/v) dilution of cDNA synthesized from the *nse4a-2* mutant. Afterwards, the PCR product was gel purified and used for the subsequent nested PCR reaction. This step was repeated until the fourth reaction. PCR product obtained from the fourth reaction was cloned into the pJET1.2 vector and sequenced.

### RNA Sequencing and Microarray Analysis

RNA for RNA sequencing was isolated using RNeasy Plant Mini kit (QIAGEN) with additional on-column DNase I digestion according to manufacturer's instructions. RNA sequencing was performed with two biological replicates per experimental point. The libraries were prepared from 1 μg of total RNA with RNA integrity number >7.8 (Bioanalyzer, Agilent) using TruSeq type RNA kit (Illumina) at the Cologne Genome Centre and sequenced as 100-bp single-end reads on a HiSeq2500 instrument (Illumina). Reads were trimmed and quality filtered with FAST-X tools (http://hannonlab.cshl.edu/fastx_toolkit/). This yielded an average of 18.5 million high-quality reads per library. The reads were mapped to the TAIR10 Arabidopsis reference genome using Tophat2 (Kim et al., 2013) with default settings. The coverage of individual genes was retrieved with the Qualimap from the set of uniquely mapped reads and significance (adjusted P-value < 0.05) of mRNA level changes estimated with the DESeq package (Anders and Huber, 2010). Publicly available *sni1-1* Affymetrix Arabidopsis ATH1 GeneChip array data (Nottingham Arabidopsis Science Centre experiment ID 389, slides 20561 to 20566; Gene Expression Omnibus Series: GSE6827; Mosher et al., 2006) were analyzed using rma protocol with Bioconductor in R. Venn diagrams were drawn using BioVenn online tool (http://www.biovenn.nl/).

### GUS Histochemical Staining

The staining protocol was adapted according to different tissues. Vegetative tissues were stained as described previously (Liu et al., 2015). Inflorescences were dissected under an MZ16FA stereomicroscope (Leica Microsystems), fixed for 30 min in ice-cold 4% (v/v) formaldehyde in 1× PBS buffer, washed three times for 5 min each in 1× PBS, and infiltrated with GUS staining solution (Stangeland and Salehian, 2002) under vacuum. After 10 to 15 min, the vacuum was released and samples were incubated at 37°C for 3 d, followed by overnight clearing in 70% (v/v) ethanol. Subsequently, inflorescences were rinsed with water and mounted in Petri dishes containing agarose and water. For staining of ovules and young seeds, developing siliques were first opened and fixed in 90% (v/v) cold acetone at −20°C for 45 min. Afterwards, they were rinsed three times with 100 mM phosphate buffer, transferred to GUS staining solution, vacuum infiltrated for 5 min, and stained at 37°C for 48 h. After staining, pistils and siliques were quickly rinsed with phosphate buffer and mounted in 8:2:1 chloral hydrate solution. In order to avoid loss of signal when we observed weak GUS staining, we performed a less severe clearing. We dissected pistils and immediately transferred them to GUS solution. Staining of ovules was performed as described previously (Vielle-Calzada et al., 2000). After clearing, mounted ovules where immediately imaged using a microscope (Zeiss). For GUS and 4′,6-diamidino-2-phenylindole (DAPI) costaining of pollen grains, flowers were opened and fixed in cold 3:1 ethanol:acetic acid (v/v) for 30 min. Afterwards, they were rinsed three times with phosphate buffer, infiltrated with GUS staining solution for 10 to 15 min, and stained for 48 h at 37°C in dark. Next, GUS-stained anthers were dissected, rinsed with phosphate buffer, transferred to a microscopic slide, further dissected with a needle in DAPI solution (0.4 μg/mL DAPI, 0.1 M sodium phosphate

buffer, pH 7, 0.1% (v/v) Triton X-100, and 1 mM EDTA), covered with a cover slip, and then used for microscopy.

### Hoyer's Clearing

Clearing of seeds was performed as described by Liu and Meinke (1998).

### Cell Cycle Arrest

The double homozygous *nse4a-2 ProCYCB1;1:CYCB1;1:GUS* and *nse4b-2 ProCYCB1;1:CYCB1;1:GUS* plants were grown for 5 d in liquid one half MS medium; transferred to liquid one half MS supplemented with 10 μM zebularine for 0, 1, 3, 6, 12, 24, or 48 h; GUS stained overnight; cleared in 70% (v/v) ethanol; and imaged using an MZ16FA stereomicroscope (Leica Microsystems).

### Confocal Microscopy

For cell death analysis, seeds from transgenic lines were grown on vertically positioned plates with one half MS medium for 4 d and then transferred for 1 d to liquid one half MS medium with 20 μM zebularine. Seedlings were stained with 10 μg mL$^{-1}$ PI solution (Sigma-Aldrich) for 3 min, followed by a rinsing step with sterilized water, and were placed on slides in a drop of water and then evaluated using an LSM700 laser scanning confocal microscope (Zeiss). For subcellular localization of NSE4A-VENUS in roots, transgenic lines expressing *ProNSE4A:VENUS:NSE4A:TerNSE4A* were grown for 5 d in either solid one half MS or one half MS supplemented with 10 μM zebularine. Afterwards, seedlings were stained with PI, and imaged with a TCS SP8 confocal microscope (Leica Microsystems). For imaging of ovules, pistils were quickly dissected in a drop of water, and ovules from different stages were mounted on a slide with a drop of water and placed on ice. After few minutes, preparations were observed using a TCS SP8 confocal microscope (Leica Microsystems)

### Y2H Assay and BiFC

The full-length CDSs of Arabidopsis *SMC5*, *SMC6A*, *SMC6B*, *NSE1*, and *NSE3* were PCR amplified from cDNA. *SMC5, SMC6A,* and *SMC6B* were cloned via restriction digest (Supplemental Table 7) into the vector *pGADT7* (Clontech), while *NSE1* and *NSE3* were cloned into the gateway compatible vector *pGADT-GW* (Lu et al., 2010) to produce a protein fusion with the GAL4 DNA activation domain (AD) in N-terminal orientation. In order to produce a protein fusion with the GAL4 DNA binding domain (BD), the *SMC5*, *NSE4A*, and *NSE4B* PCR fragments were cloned via restriction (Supplemental Table 7) digest into *pGBKT7* and *NSE1* and *NSE3* were cloned via gateway into *pGBKT7-GW* (Lu et al., 2010). In order to avoid negative results due to interference of BD or AD domain with possible interactors, all genes were cloned into both C-terminal *pGBKCg* and *pGADCg* Y2H vectors, to produce C-terminally tagged GAL4 AD and BD fusion proteins, respectively, with exception of NSE4B, which was only cloned into the *pGADCg* vector. The hinge and fragments of coils of SMC5 (corresponding to amino acids 415 to 699), SMC6A (amino acids 367 to 670), and SMC6B (amino acids 358 to 691) were cloned into the *pGBKCg* and *pGADCg* vectors to test for interaction with the core subunits. The GAL4-based interaction was tested in the yeast strain AH109 (Clontech). Cotransformed yeast strains were selected on synthetic defined/–Leu/–Trp medium. Protein–protein interactions were tested using stringent (synthetic defined/–Leu/–Trp/–His) selection medium supplemented with defined concentrations of 3-AT (Supplemental Table 4). The interaction between *pGADT7-T* and *pBKT7-53* was used as the positive control and that between *pGADT7-T* and *pBKT7-LamC* was used as the negative control. For BiFC, we used the same CDSs as for the Y2H experiments. The *SMC5, SMC5* (hinge), and *NSE3* sequences were cloned into *pBATL-*

*nYFP*, and *NSE4A*, *NSE4B*, *NSE1*, and *SMC6B* hinge sequences were cloned into *pBaTL-cYFP*. Both plasmids produce C-terminal fusion proteins. *Nicotiana benthamiana* leaves were transformed for transient expression as described previously (Tian et al., 2011). YFP fluorescence was observed using an LSM 700 confocal microscope (Zeiss).

### Coimmunoprecipitation and Localization Assays

Here, we used the same entry clones as for Y2H and BiFC assays. Hinge *SMC5* was cloned into *pGWB560*, to produce a C-terminal fusion with tagRFP protein. Hinge *SMC6A*, hinge *SMC6B*, *NSE1* CDS, *NSE4A* CDS, and *NSE4B* CDS were cloned into *pGWB541* to produce a C-terminal tagged EYFP proteins. *NSE3* CDS was cloned into *pGWB611* to produce a C-terminal FLAG fusion protein. To test interactions of SMC5 with NSE4s, full-length *SMC5* CDS was cloned into *pGWB561* to produce an N-terminal tagRFP fusion, while both *NSE4A* and *NSE4B* CDS were cloned into *pGWB542* to produce N- and C-terminal tagged EYFP fusion proteins. As a negative control we used *pSY1*, containing GFP CDS driven by 35S promoter. Afterwards, the expression clones were transformed into *A. tumefaciens* strain GV3101.

Fluorescent or epitope tag–conjugated proteins were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration. Leaves were harvested at 4 or 5 d after inoculation, and immunoprecipitation was performed with a μMACS GFP isolation kit (Miltenyi Biotec). Approximately 1 to 2 g of plant material was homogenized in threefold volume of μMACS lysis buffer containing protease inhibitor cocktail for plant cell and tissue extracts (Sigma-Aldrich), and then the lysate was filtered through two layers of miracloth. Afterwards, the lysate was mixed with anti-GFP antibody-conjugated magnetic beads and was incubated at 4°C for 60 min with gentle rotation. The GFP-conjugated proteins were purified using a magnetic column according to the manufacturer's instructions. The immunoprecipitated proteins were analyzed by protein gel blotting using an anti-GFP antibody at 1/1000 (v/v; ab290, Abcam), an anti-tagRFP antibody at 1/500 (v/v; R10367, Thermo Fisher Scientific), or an anti-FLAG antibody at 1/5000 (v/v; 3022-100, BioVision) as primary antibodies and horseradish peroxidase–conjugated anti-mouse IgG antibody at 1/15000 (v/v; W402, Promega) or horseradish peroxidase–conjugated anti-rabbit IgG antibody at 1/15000 (v/v; MB4458, MBL) as secondary antibodies. The chemiluminescences from target proteins of each antibody were visualized with ImmunoStar LD (Wako) on Fusion Pulse system (Vilber Lourmat).

For the localization analysis of GFP, EYFP-NSE4A, and EYFP-NSE4B proteins simultaneously expressed with tagRFP-SMC5. Five days after inoculation, leaves were observed under an inverted FV1200 laser confocal microscope equipped with a GaAsP detector (Olympus) with an excitation wavelength with 473 nm for GFP/EYFP and 559 nm for tagRFP.

### Statistical Analysis

The values were examined by one-way analysis of variance and post hoc comparison by Duncan's multiple range test ($P \leq 0.05$). Statistical analyses except for RNA sequencing and microarray analysis were performed using STATISTICA 13 software (StatSoft). Fisher's test was used to calculate the adjusted P-value (q-value) in RNA sequencing and microarray analysis. Raw data and detailed results of the statistical analyses are provided in Supplemental Data Set 6.

### Accession Numbers

The following gene names and symbols are associated with this publication: *ASAP1* (AT2G28130), *ATR* (AT5G40820), *LIG4* (AT5G57160), *HPY2* (AT3G15150), *NSE1* (AT5G21140), *NSE3* (AT1G34770), *NSE4A* (AT1G51130), *NSE4B* (AT3G20760), *SMC5* (AT5G15920), *SMC6A* (AT5G07660), *SMC6B* (AT5G61460), *SNI1* (AT4G18470), *WEE1*

(AT1G02970). RNA sequencing reads are deposited in Gene Expression Omnibus as the study number GSE113310.

### Supplemental Data

**Supplemental Figure 1.** Phylogenetic shadowing of *NSE4A* and *NSE4B* promoters.

**Supplemental Figure 2.** Characterization of *NSE4A* mutation in *nse4a-2*.

**Supplemental Figure 3.** Characterization of *nse4b-1* and *nse4b-2* mutations.

**Supplemental Figure 4.** Chromatin environment of the *NSE4A* genomic region.

**Supplemental Figure 5.** Chromatin environment of the *NSE4B* genomic region.

**Supplemental Figure 6.** Seed phenotypes of *NSE4A* mutants.

**Supplemental Figure 7.** Mutant sensitivity to hydroxyurea (HU).

**Supplemental Figure 8.** Y2H assays (autoactivation and negative results).

**Supplemental Figure 9.** Whole blots from coimmunoprecipitation (co-IP) assays.

**Supplemental Table 1.** Protein sequences used in the phylogenetic analysis.

**Supplemental Table 2.** Promoter sequences used for the phylogenetic shadowing.

**Supplemental Table 3.** Gene ontology (GO) term analysis of 82 genes upregulated in *nse4a-2* and *sni1-1*.

**Supplemental Table 4.** 3-AT concentrations used in yeast-two hybrid experiments.

**Supplemental Table 5.** Punnett square indicating frequencies of genotypes in F2 generation of self-pollinated F1 hybrid *nse4a-1/ nse4a-2 T/0*.

**Supplemental Table 6.** Promoter swap rescues *nse4a-1* lethality.

**Supplemental Table 7.** Primers used in this study.

**Supplemental Table 8.** Primer combinations used in 3′ RACE PCR of *NSE4A* in *nse4a-2*.

**Supplemental Data Set 1.** Protein sequences used to build NSE4 phylogenetic tree (FASTA file).

**Supplemental Data Set 2.** Alignment from the full length NSE4 protein sequences submitted to Gblocks server (FASTA file).

**Supplemental Data Set 3.** Gblocks output. Conserved blocks from NSE4 protein alignment (FASTA file).

**Supplemental Data Set 4.** Transcriptomic study of *nse4a-2* under control and zebularine stress conditions.

**Supplemental Data Set 5.** Comparison of *nse4a-2* and *sni1-1* induced transcriptional changes.

**Supplemental Data Set 6.** Background data for and the results of the statistical analyses.

## AUTHOR CONTRIBUTIONS

A.P. and M.D. designed the research with help from C.B., T.S., U.G., and S.M. on specific aspects. M.D., P.P., A.N., T.S., H.J., C.B., and P.Y.G. performed the experiments and analyzed the data. A.P. and M.D. wrote the article with contributions from all authors. All authors approved the final version of this article.

## REFERENCES

**Abdelsamad, A., and Pecinka, A.** (2014). Pollen-specific activation of Arabidopsis retrogenes is associated with global transcriptional reprogramming. Plant Cell **26:** 3299–3313.

**Alabert, C., and Groth, A.** (2012). Chromatin replication and epigenome maintenance. Nat. Rev. Mol. Cell Biol. **13:** 153–167.

**Anders, S., and Huber, W.** (2010). Differential expression analysis for sequence count data. Genome Biol. **11:** R106.

**Awasthi, P., Foiani, M., and Kumar, A.** (2015). ATM and ATR signaling at a glance. J. Cell Sci. **128:** 4255–4262.

**Baroux, C., and Autran, D.** (2015). Chromatin dynamics during cellular differentiation in the female reproductive lineage of flowering plants. Plant J. **83:** 160–176.

**Baroux, C., Pecinka, A., Fuchs, J., Schubert, I., and Grossniklaus, U.** (2007). The triploid endosperm genome of Arabidopsis adopts a peculiar, parental-dosage-dependent chromatin organization. Plant Cell **19:** 1782–1794.

**Bowman, J.L., Mansfield, S.G., Modrusan, Z., Reiser, L., Fischer, R.L., Haughn, G.W., Feldman, K.A., and Webb, M.C.** (1994). Ovules. In Arabidopsis, J. Bowman, ed (New York: Springer), pp. 297–331.

**Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., Batzoglou, S., and Batzoglou, S.; NISC Comparative Sequencing Program**. (2003). LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. **13:** 721–731.

**Castresana, J.** (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17:** 540–552.

**Chiolo, I., Minoda, A., Colmenares, S.U., Polyzos, A., Costes, S.V., and Karpen, G.H.** (2011). Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. Cell **144:** 732–744.

**Colón-Carmona, A., You, R., Haimovitch-Gal, T., and Doerner, P.** (1999). Technical advance: Spatio-temporal analysis of mitotic activity with a labile cyclin-GUS fusion protein. Plant J. **20:** 503–508.

**De Piccoli, G., Torres-Rosell, J., and Aragón, L.** (2009). The unnamed complex: What do we know about Smc5-Smc6? Chromosome Res. **17:** 251–263.

**De Schutter, K., Joubès, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., Van Der Schueren, E., Beeckman, T., Kushnir, S., Inzé, D., and De Veylder, L.** (2007). Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. Plant Cell **19:** 211–225.

**Diaz, M., and Pecinka, A.** (2017). Seeds as emerging hotspot for maintenance of genome stability. Cytologia (Tokyo) **82:** 467–480.

**Diaz, M., and Pecinka, A.** (2018). Scaffolding for repair: Understanding molecular functions of the SMC5/6 complex. Genes (Basel) **9:** 36.

**Duan, X., Yang, Y., Chen, Y.-H., Arenz, J., Rangi, G.K., Zhao, X., and Ye, H.** (2009). Architecture of the Smc5/6 complex of *Saccharomyces cerevisiae* reveals a unique interaction between the Nse5–6 subcomplex and the hinge regions of Smc5 and Smc6. J. Biol. Chem. **284:** 8507–8515.

**Edgar, R.C.** (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:** 1792–1797.

**Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I.** (2004). VISTA: Computational tools for comparative genomics. Nucleic Acids Res. **32:** W273-279.

**Hanin, M., Mengiste, T., Bogucki, A., and Paszkowski, J.** (2000). Elevated levels of intrachromosomal homologous recombination in Arabidopsis overexpressing the MIM gene. Plant J. **24:** 183–189.

**Hirano, T.** (2006). At the heart of the chromosome: SMC proteins in action. Nat. Rev. Mol. Cell Biol. **7:** 311–322.

**Hirano, T.** (2012). Condensins: Universal organizers of chromosomes with diverse functions. Genes Dev. **26:** 1659–1678.

**Huang, L., Yang, S., Zhang, S., Liu, M., Lai, J., Qi, Y., Shi, S., Wang, J., Wang, Y., Xie, Q., and Yang, C.** (2009). The Arabidopsis SUMO E3 ligase AtMMS21, a homologue of NSE2/MMS21, regulates cell proliferation in the root. Plant J. **60:** 666–678.

**Hudson, J.J.R., Bednarova, K., Kozakova, L., Liao, C., Guerineau, M., Colnaghi, R., Vidot, S., Marek, J., Bathula, S.R., Lehmann, A.R., and Palecek, J.** (2011). Interactions between the Nse3 and Nse4 components of the SMC5-6 complex identify evolutionarily conserved interactions between MAGE and EID Families. PLoS One **6:** e17270.

**Ishida, T., Fujiwara, S., Miura, K., Stacey, N., Yoshimura, M., Schneider, K., Adachi, S., Minamisawa, K., Umeda, M., and Sugimoto, K.** (2009). SUMO E3 ligase HIGH PLOIDY2 regulates endocycle onset and meristem maintenance in Arabidopsis. Plant Cell **21:** 2284–2297.

**Jeppsson, K., Carlborg, K.K., Nakato, R., Berta, D.G., Lilienthal, I., Kanno, T., Lindqvist, A., Brink, M.C., Dantuma, N.P., Katou, Y., Shirahige, K., and Sjögren, C.** (2014a). The chromosomal association of the Smc5/6 complex depends on cohesion and predicts the level of sister chromatid entanglement. PLoS Genet. **10:** e1004680.

**Jeppsson, K., Kanno, T., Shirahige, K., and Sjögren, C.** (2014b). The maintenance of chromosome structure: Positioning and functioning of SMC complexes. Nat. Rev. Mol. Cell Biol. **15:** 601–614.

**Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., Links, M.G., Sharpe, A.G., and Parkin, I.A.P.** (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell **26:** 2777–2791.

**Kanno, T., Berta, D.G., and Sjögren, C.** (2015). The Smc5/6 complex is an ATP-dependent intermolecular DNA linker. Cell Reports **12:** 1471–1482.

**Kegel, A., and Sjögren, C.** (2010). The Smc5/6 complex: More than repair? Cold Spring Harb. Symp. Quant. Biol. **75:** 179–187.

**Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L.** (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14:** R36.

**Kozak, J., West, C.E., White, C., da Costa-Nunes, J.A., and Angelis, K.J.** (2009). Rapid repair of DNA double strand breaks in *Arabidopsis thaliana* is dependent on proteins involved in chromosome structure maintenance. DNA Repair (Amst.) **8:** 413–419.

**Kwak, J.S., Son, G.H., Kim, S.-I., Song, J.T., and Seo, H.S.** (2016). Arabidopsis HIGH PLOIDY2 sumoylates and stabilizes Flowering Locus C through its E3 ligase activity. Front. Plant Sci. **7:** 530.

**Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J.U., and Forner, J.** (2013). GreenGate---A novel, versatile, and efficient cloning system for plant transgenesis. PLoS One **8:** e83043.

**Li, G., Zou, W., Jian, L., Qian, J., Deng, Y., and Zhao, J.** (2017). Non-SMC elements 1 and 3 are required for early embryo and seedling development in Arabidopsis. J. Exp. Bot. **68:** 1039–1054.

**Li, X., Zhang, Y., Clarke, J.D., Li, Y., and Dong, X.** (1999). Identification and cloning of a negative regulator of systemic acquired resistance, SNI1, through a screen for suppressors of npr1-1. Cell **98:** 329–339.

**Liu, C.-M., and Meinke, D.W.** (1998). The titan mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. Plant J. **16:** 21–31.

**Liu, C.M., McElver, J., Tzafrir, I., Joosen, R., Wittich, P., Patton, D., Van Lammeren, A.A.M., and Meinke, D.** (2002). Condensin and cohesin knockouts in Arabidopsis exhibit a titan seed phenotype. Plant J. **29:** 405–415.

**Liu, C.-H., Finke, A., Díaz, M., Rozhon, W., Poppenberger, B., Baubec, T., and Pecinka, A.** (2015). Repair of DNA damage induced by the cytidine analog zebularine requires ATR and ATM in Arabidopsis. Plant Cell **27:** 1788–1800.

**Liu, M., Shi, S., Zhang, S., Xu, P., Lai, J., Liu, Y., Yuan, D., Wang, Y., Du, J., and Yang, C.** (2014). SUMO E3 ligase AtMMS21 is required for normal meiosis and gametophyte development in Arabidopsis. BMC Plant Biol. **14:** 153.

**Lu, Q., Tang, X., Tian, G., Wang, F., Liu, K., Nguyen, V., Kohalmi, S.E., Keller, W.A., Tsang, E.W.T., Harada, J.J., Rothstein, S.J., and Cui, Y.** (2010). Arabidopsis homolog of the yeast TREX-2 mRNA export complex: Components and anchoring nucleoporin. Plant J. **61:** 259–270.

**Mengiste, T., Revenkova, E., Bechtold, N., and Paszkowski, J.** (1999). An SMC-like protein is required for efficient homologous recombination in Arabidopsis. EMBO J. **18:** 4505–4512.

**Menolfi, D., Delamarre, A., Lengronne, A., Pasero, P., and Branzei, D.** (2015). Essential roles of the Smc5/6 complex in replication through natural pausing sites and endogenous DNA damage tolerance. Mol. Cell **60:** 835–846.

**Mosher, R.A., Durrant, W.E., Wang, D., Song, J., and Dong, X.** (2006). A comprehensive structure-function analysis of Arabidopsis SNI1 defines essential regions and transcriptional repressor activity. Plant Cell **18:** 1750–1765.

**Mozgova, I., and Hennig, L.** (2015). The polycomb group protein regulatory network. Annu. Rev. Plant Biol. **66:** 269–296.

**Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A.** (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat. Biotechnol. **20:** 87–90.

**Palecek, J.J., and Gruber, S.** (2015). Kite proteins: A superfamily of SMC/kleisin partners conserved across bacteria, archaea, and eukaryotes. Structure **23:** 2183–2190.

**Palecek, J., Vidot, S., Feng, M., Doherty, A.J., and Lehmann, A.R.** (2006). The Smc5-Smc6 DNA repair complex. bridging of the Smc5-Smc6 heads by the KLEISIN, Nse4, and non-Kleisin subunits. J. Biol. Chem. **281:** 36952–36959.

**Paterson, A.H., et al.** (2009). The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551–556.

**Pebernard, S., Wohlschlegel, J., McDonald, W.H., Yates III, J.R., and Boddy, M.N.** (2006). The Nse5-Nse6 dimer mediates DNA repair roles of the Smc5-Smc6 complex. Mol. Cell. Biol. **26:** 1617–1630.

**Potts, P.R., and Yu, H.** (2007). The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. Nat. Struct. Mol. Biol. **14:** 581–590.

**Puchta, H., Swoboda, P., and Hohn, B.** (1995). Induction of intrachromosomal homologous recombination in whole plants. Plant J. **7:** 203–210.

**Räschle, M., et al.** (2015). DNA repair. Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. Science **348:** 1253671.

**Schoft, V.K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H.** (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep. **10:** 1015–1021.

**Schubert, V.** (2009). SMC proteins and their multiple functions in higher plants. Cytogenet. Genome Res. **124:** 202–214.

**Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A.** (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136:** 461–472.

**Stangeland, B., and Salehian, Z.** (2002). An improved clearing method for GUS assay in Arabidopsis endosperm and seeds. Plant Mol. Biol. Report. **20:** 107–114.

**Stingele, J., and Jentsch, S.** (2015). DNA-protein crosslink repair. Nat. Rev. Mol. Cell Biol. **16:** 455–460.

**Tian, G., Lu, Q., Zhang, L., Kohalmi, S.E., and Cui, Y.** (2011). Detection of protein interactions in plant using a gateway compatible bimolecular fluorescence complementation (BiFC) system. J. Vis. Exp. **55:** 3473.

**Torres-Rosell, J., Sunjevaric, I., De Piccoli, G., Sacher, M., Eckert-Boulet, N., Reid, R., Jentsch, S., Rothstein, R., Aragón, L., and Lisby, M.** (2007). The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. Nat. Cell Biol. **9:** 923–931.

**Tretyakova, N.Y., Groehler IV, A., and Ji, S.** (2015). DNA-protein cross-links: Formation, structural identities, and biological outcomes. Acc. Chem. Res. **48:** 1631–1644.

**Tzafrir, I., McElver, J.A., Liu Cm, C.M., Yang, L.J., Wu, J.Q., Martinez, A., Patton, D.A., and Meinke, D.W.** (2002). Diversity of TITAN functions in Arabidopsis seed development. Plant Physiol. **128:** 38–51.

**Uhlmann, F.** (2016). SMC complexes: From DNA to chromosomes. Nat. Rev. Mol. Cell Biol. **17:** 399–412.

**Vielle-Calzada, J.-P., Baskar, R., and Grossniklaus, U.** (2000). Delayed activation of the paternal genome during seed development. Nature **404:** 91–94.

**Watanabe, K., Pacher, M., Dukowic, S., Schubert, V., Puchta, H., and Schubert, I.** (2009). The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 complex promotes sister chromatid alignment and homologous recombination after DNA damage in *Arabidopsis thaliana*. Plant Cell **21:** 2688–2699.

**Wu, N., and Yu, H.** (2012). The Smc complexes in DNA damage response. Cell Biosci. **2:** 5.

**Xu, P., Yuan, D., Liu, M., Li, C., Liu, Y., Zhang, S., Yao, N., and Yang, C.** (2013). AtMMS21, an SMC5/6 complex subunit, is

involved in stem cell niche maintenance and DNA damage responses in Arabidopsis roots. Plant Physiol. **161:** 1755–1768.

**Yan, S., Wang, W., Marqués, J., Mohan, R., Saleh, A., Durrant, W.E., Song, J., and Dong, X.** (2013). Salicylic acid activates DNA damage responses to potentiate plant immunity. Mol. Cell **52:** 602–610.

**Yuan, D., Lai, J., Xu, P., Zhang, S., Zhang, J., Li, C., Wang, Y., Du, J., Liu, Y., and Yang, C.** (2014). AtMMS21 regulates DNA damage response and homologous recombination repair in Arabidopsis. DNA Repair (Amst.) **21:** 140–147.

**Zabrady, K., Adamus, M., Vondrova, L., Liao, C., Skoupilova, H., Novakova, M., Jurcisinova, L., Alt, A., Oliver, A.W., Lehmann, A.R., and Palecek, J.J.** (2016). Chromatin association of the SMC5/6 complex is dependent on binding of its NSE3 subunit to DNA. Nucleic Acids Res. **44:** 1064–1079.

**Zhang, S., Qi, Y., Liu, M., and Yang, C.** (2013). SUMO E3 ligase AtMMS21 regulates drought tolerance in Arabidopsis thaliana(F). J. Integr. Plant Biol. **55:** 83–95.

**Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H.** (2006). *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. Nat. Protoc. **1:** 641–646.

**Zhao, X., and Blobel, G.** (2005). A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. Proc. Natl. Acad. Sci. USA **102:** 4777–4782.

Mariana Díaz, Petra Pecinková, Anna Nowicka, Célia Baroux, Takuya Sakamoto, Priscilla Yuliani Gandha, Hana Jerábková, Sachihiro Matsunaga, Ueli Grossniklaus and Ales Pecinka

This information is current as of July 10, 2019

| | |
|---|---|
| **Supplemental Data** | /content/suppl/2019/04/29/tpc.18.00043.DC1.html<br>/content/suppl/2019/05/18/tpc.18.00043.DC2.html |
| **References** | This article cites 71 articles, 21 of which can be accessed free at:<br>/content/31/7/1579.full.html#ref-list-1 |
| **Permissions** | https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X |
| **eTOCs** | Sign up for eTOCs at:<br>http://www.plantcell.org/cgi/alerts/ctmain |
| **CiteTrack Alerts** | Sign up for CiteTrack Alerts at:<br>http://www.plantcell.org/cgi/alerts/ctmain |
| **Subscription Information** | Subscription Information for *The Plant Cell* and *Plant Physiology* is available at:<br>http://www.aspb.org/publications/subscriptions.cfm |

# Publikace 4

# Genome invasion by a hypomethylated satellite repeat in Australian crucifer *Ballantinia antipoda*

Andreas Finke[1], Terezie Mandáková[2], Kashif Nawaz[1,3], Giang T. H. Vu[1,†], Petr Novák[4], Jiri Macas[4], Martin A. Lysak[2] and Ales Pecinka[1,3,*] (ID)

[1]*Max Planck Institute for Plant Breeding Research (MPIPZ), Cologne 50829, Germany,*

[2]*Plant Cytogenomics Research Group, CEITEC – Central-European Institute of Technology, Masaryk University, Brno 62500, Czech Republic,*

[3]*The Czech Academy of Sciences, Institute of Experimental Botany (IEB), Centre of the Region Haná for Agricultural and Biotechnological Research (CRH), Olomouc 77900, Czech Republic, and*

[4]*Biology Centre, The Czech Academy of Sciences, České Budejovice 37005, Czech Republic*

## SUMMARY

**Repetitive sequences are ubiquitous components of all eukaryotic genomes. They contribute to genome evolution and the regulation of gene transcription. However, the uncontrolled activity of repetitive sequences can negatively affect genome functions and stability. Therefore, repetitive DNAs are embedded in a highly repressive heterochromatic environment in plant cell nuclei. Here, we analyzed the sequence, composition and the epigenetic makeup of peculiar non-pericentromeric heterochromatic segments in the genome of the Australian crucifer *Ballantinia antipoda*. By the combination of high throughput sequencing, graph-based clustering and cytogenetics, we found that the heterochromatic segments consist of a mixture of unique sequences and an A−T-rich 174 bp satellite repeat (*BaSAT1*). *BaSAT1* occupies about 10% of the *B. antipoda* nuclear genome in >250 000 copies. Unlike many other highly repetitive sequences, *BaSAT1* repeats are hypomethylated; this contrasts with the normal patterns of DNA methylation in the *B. antipoda* genome. Detailed analysis of several copies revealed that these non-methylated *BaSAT1* repeats were also devoid of heterochromatic histone H3K9me2 methylation. However, the factors decisive for the methylation status of *BaSAT1* repeats remain currently unknown. In summary, we show that even highly repetitive sequences can exist as hypomethylated in the plant nuclear genome.**

Keywords: satellite repeats, heterochromatin, DNA methylation, comparative genomics, *Brassicaceae*.

## INTRODUCTION

Repetitive sequences, including transposable elements (TEs) and satellite repeats, are ubiquitous components of eukaryotic genomes and have major effects on genome organization, evolution and gene regulation (Lisch, 2013; Mehrotra and Goyal, 2014). In flowering plants, repetitive DNA content varies from less than 10% in miniature genomes of highly specialized carnivorous plants *Utricularia gibba* and *Genlisea nigrocaulis* to 85% in maize (Schnable *et al.*, 2009; Ibarra-Laclette *et al.*, 2013; Vu *et al.*, 2015). The full spectrum and interplay of factors determining repetitive DNA content per genome remain unknown and represent part of the C-value enigma (Gregory, 2005). Many TEs produce their own proteins necessary for amplification, and particularly autonomous RNA transposons

(retrotransposons), multiplying via a copy−paste mechanism, have been very successful in invading plant genomes ove short evolutionary times (Piegu *et al.*, 2006; Willing *et al.*, 2015). Recent studies have suggested that retrotransposons contain *cis*-regulatory sequences that are recognized by specific transcription factors and therefore link TE expression with the canonical gene regulatory pathways (Ito *et al.*, 2011; Cavrak *et al.*, 2014; Pietzenuk *et al.*, 2016). In contrast with TEs, which are often several kilobases long and dispersed in the genome, satellite DNAs form homogenous, up to mega base pair long, arrays consisting of a high copy number of typically shorter (150–400 bp) sequence motifs (Heslop-Harrison and Schwarzacher, 2011; Melters *et al.*, 2013; Garrido-Ramos,

2015). The distribution of satellite repeats varies along chromosomes. While the chromosome starts and ends with telomeric repeats, the position of other regions with high density of satellite repeats, including centromeres, nucleolar organizers (NORs) and heterochromatic knobs in some species, for example, maize (Gent *et al.*, 2014), is variable (Mehrotra and Goyal, 2014; Garrido-Ramos, 2015). With exception of ribosomal and telomeric repeats, satellite DNAs are less conserved and mostly specific for a single or few closely related species. The origin of satellites is not yet fully understood, but it has been shown that they can arise *de novo* or by amplification of short tandem repeat arrays already present in the genome as parts of retrotransposons or rDNA ITS sequences (Macas *et al.*, 2003, 2009). Satellite repeats are most likely to amplify via the combinatorial action of unequal crossing over, gene conversion, rolling circle amplification and/or replication slippage (Plohl *et al.*, 2008; Garrido-Ramos, 2015). Some satellite DNAs have essential functions including protection of chromosome ends by telomeres, acting as a platform for kinetochore binding by centromeres or producing high amounts of ribosomal RNA by NORs (Mehrotra and Goyal, 2014). Specialized satellite functions include the regulation of gene expression or effects on chromosome segregation via meiotic drive (Belele *et al.*, 2013; Dawe *et al.*, 2018). Another important function of satellites and other repeats is by creating sequence diversity, which accelerates formation of reproductive barriers (Garrido-Ramos, 2015).

Amplification of TEs, is opposed both epigenetically and genetically. In epigenetic silencing, repeat-derived transcripts are processed into small RNAs, this devalues them as templates for reverse transcription (Mari-Ordonez *et al.*, 2013) and guides the RNA-directed DNA methylation (RdDM) machinery to homologous sequences (reviewed in e.g. Matzke and Mosher, 2014). These regions will be DNA methylated in CG, CHG and CHH contexts (H = A, T or C), histone H3 lysine 9 dimethylated (H3K9me2) and transcriptionally repressed. The given epigenetic state is faithfully transmitted to the next generations and remains robust under various growth situations due to the meristematic silencing centers (Yadav *et al.*, 2009; Du *et al.*, 2012; Baubec *et al.*, 2014). At the same time, TEs are subject to fast mutagenesis via the deamination of methyl-cytosines, microdeletions and deletion-prone homologous recombination events (Devos *et al.*, 2002; Hawkins *et al.*, 2009; Hu *et al.*, 2011; Willing *et al.*, 2016). Although we assume that similar mechanisms control satellite repeats, it is yet to be elucidated if, and how, their proliferation is regulated and eventually suppressed. Data from maize suggest that satellite repeat arrays are less targeted by RdDM at least during vegetative development under ambient conditions (Gent *et al.*, 2014; Fu *et al.*, 2018).

While the distribution of repeats along chromosomes is variable, several common patterns can be observed among plant genomes. In species with small genomes (<500 Mbp/1C) and low repeat content, repetitive DNA forms typically a single major chromosomal cluster containing the centromeric satellite array flanked by pericentromeric regions rich in various TEs and satellite DNA (Ali *et al.*, 2005; Mandáková *et al.*, 2010; Seymour *et al.*, 2014; Vu *et al.*, 2015; Willing *et al.*, 2015). In addition, some species also form repeat-rich domains at the chromosome termini. (Peri)centromeric repeats and inactive rDNA repeats form compact microscopically visible nuclear membrane- or nucleolus-associated heterochromatic chromocenters (CCs), respectively (Fransz *et al.*, 2003, 2006). Increasing genome size, is usually associated with the presence of repeats in chromosome arms. In plants with small genomes, chromosomes often adopt a rosette-like organization during interphase (Fransz *et al.*, 2002), while in plants with large genomes they appear heterochromatic and are often organized with centromeres and telomeres clustered at opposite nuclear poles (Cowan *et al.*, 2001; Tiang *et al.*, 2012). Genomes of crucifers (*Brassicaceae*) with small genomes show the first type of heterochromatin distribution with minor differences caused by the presence of, for example, heterochromatic knobs (Lysak *et al.*, 2005; Mandáková and Lysak, 2008; Hay *et al.*, 2014; Fransz *et al.*, 2016). A remarkable exception in this pattern was found in the endemic Australian species *Ballantinia antipoda* (*B. antipoda*; Southern Shepherd's Purse) with a small genome ($2n = 12$; 1C ~472 Mbp), but with six heterochromatic segments (HSs) occupying up to the entire chromosome arm (Mandáková *et al.*, 2010; Majerová *et al.*, 2014) (Figure 1a).

## RESULTS

### A 174-bp satellite repeat is a principal component of HSs on *B. antipoda* chromosomes

We hypothesized that the HSs on *B. antipoda* chromosomes are formed by a specific highly amplified repeat. Therefore, we investigated the most abundant repetitive DNA by analyzing *B. antipoda* genomic shotgun reads using a RepeatExplorer pipeline (Novák *et al.*, 2013). The pipeline performs all-to-all pairwise similarity comparisons of sequence reads and identifies genomic repeats as clusters of frequently overlapping read sequences (Novák *et al.*, 2010). Clustering of 865 000 randomly sampled reads ($\approx 0.2\times$ the nuclear genome) resulted in 1000s of clusters ranging from two up to 71 000 reads, and therefore reflecting varying abundance of corresponding genomic sequences. We found 89 clusters, each containing at least 0.05% of the analyzed reads, that were considered to represent abundant repeats. The clusters corresponded to 49.2% of *B. antipoda* nuclear genome and were mostly
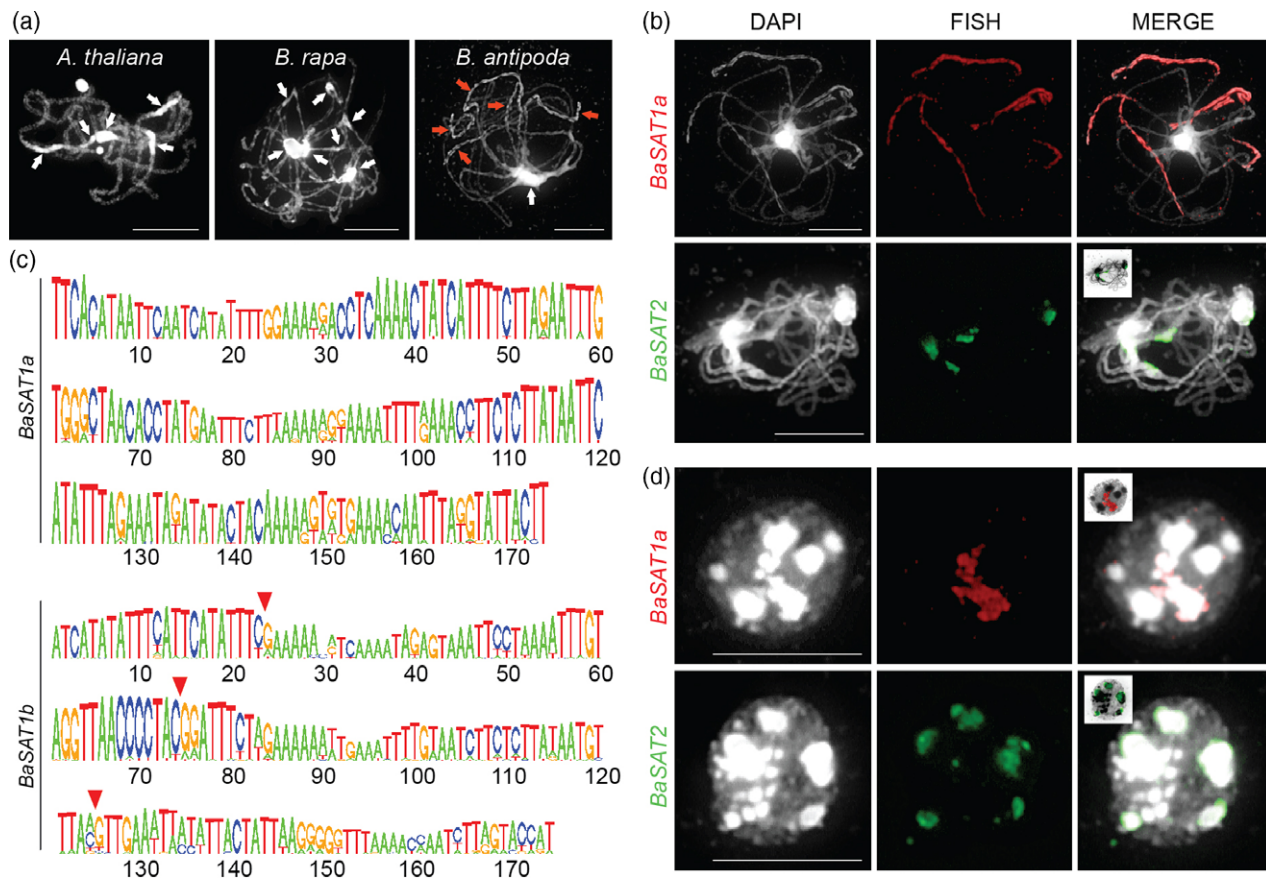
**Figure 1.** Localization and composition of heterochromatic segments in *B. antipoda*.
(a) Pachytene chromosomes of *Arabidopsis thaliana*, *Brassica rapa* and *B. antipoda* with indicated pericentromeric heterochromatin (white arrows) and *B. antipoda* heterochromatic segments (red arrows).
(b) FISH on *B. antipoda* pachytene chromosomes using probes against *BaSAT1a* (red) and *BaSAT2* (green). The probes had 95% and 94% sequence identity with the *BaSAT1a* and *BaSAT2* consensus sequences, respectively.
(c) Logo plot of the 174-bp monomer consensus sequences of the *BaSAT1a* and *BaSAT1b* repeats. Three CG dinucleotides in *BaSAT1b* are indicated by red triangles.
(d) FISH on *B. antipoda* interphase nuclei using probes against *BaSAT1a* (red) and *BaSAT2* (green). For more images see also Figure S4. All preparations in (a), (b) and (d) were counterstained with DAPI. Scale bars = 10 μm.

composed of LTR-retrotransposons and satellite DNA (16.5% and 15.9% of the genome, respectively, Table 1).

The two major satellite DNA families representing the primary candidates for the HS repeats were named *BaSAT1* and *BaSAT2* (Figures S1 and S2). These subfamilies were split into separate clusters in the RepeatExplorer analysis due to their sequence divergence (66% identity between the consensus sequences). The *BaSAT2* (5.65% of the genome) was composed of about 600-bp long monomers, which contained short arrays of partially degenerated telomeric motifs (TTTAGGG) (Figure S2). Fluorescence *in situ* hybridization (FISH) on the extended meiotic chromosomes using a *BaSAT2* specific probe labeled the middle regions of all six *B. antipoda* chromosomes (Figure 1b), suggesting that it is a (peri)centromeric repeat. The *BaSAT1*, with 174-bp monomers, comprised two distinct subfamilies designated as *BaSAT1a* and *BaSAT1b* (Figures 1c and S1a). The *BaSAT1a/b* probe

**Table 1** Composition of the highly repetitive fraction of the *B. antipoda* genome. 'All' indicates the sum of a given repeat type within *B. antipoda* genome according to graph-based clustering

| Repeat | Genome proportion (%) |
|---|---|
| Satellites (all) | (15.85) |
|   BaSAT1a | 8.19 |
|   BaSAT1b | 1.94 |
|   BaSAT2 | 5.65 |
| LTR-retrotransposons (all) | (16.47) |
| LTR/gypsy | |
|   Athila | 9.70 |
|   Chromovirus | 1.94 |
| LTR/copia | 0.99 |
| LTR/unclassified | 3.84 |
| DNA transposons (all) | (2.12) |
|   Mutator | 1.13 |
|   CACTA | 0.99 |
| rDNA | 3.55 |
| Unclassified repeats | 11.23 |
| Total | 49.23 |

unambiguously labeled all six HSs (Figure 1b) and confirmed these repeats as the principal components of HSs. During interphase, *BaSAT1* formed a high number of mini-chromocenters (CCs) without any obvious peripheral localization (Figure 1d; see also Figure S4 interphase nuclei). Considering estimated genome proportions of *BaSAT1* repeats, their prevailing monomer length and haploid genome size (~472 Mbp), we estimated genomic copy numbers of *BaSAT1a* and *BaSAT1b* repeats to be approximately 212 000 and 50 000, respectively. The subfamilies made up 8.19% and 1.94% of the nuclear genome, respectively, making *BaSAT1* the most abundant repeat in *B. antipoda*. Detailed analysis of the *BaSAT1a* and *BaSAT1b* consensus sequences revealed that they are very A−T rich (76.4 and 75.3%; Figures 1b and S1). All cytosines in the *BaSAT1a* consensus sequence were in the CHH (H = C, A or T) context, while the *BaSAT1b* consensus sequence also contained three CG dinucleotides (Figures 1c and S1a).

### *BaSAT1* repeats are distributed in gene-rich chromosome regions

Comparative chromosome painting using *A. thaliana* gene-rich BAC probes revealed their hybridization to *B. antipoda* HSs (Mandáková *et al.*, 2010), indicating that HSs also contain single copy sequences. To get further insight into the organization of HSs, we combined the *BaSAT1* FISH probe with distinctly labeled *A. thaliana* BAC FISH probes from the bottom arm of chromosome 2 (evolutionary conserved block J), which mark homeologous regions on *B. antipoda* chromosomes 3 and 6 (Mandáková *et al.*, 2010), and hybridized them to *B. antipoda* pachytene chromosomes. Indeed, the BAC and *BaSAT1* signals alternated in a mosaic, proving that HSs contain a mixture of repetitive and evolutionary conserved single copy sequences (Figure 2a). We identified part of these sequences and their organization relative to *BaSAT1* repeats by high-throughput sequencing and *de novo* contig assembly of 115 million *B. antipoda* 100-bp single-end Illumina (San Diego, CA, USA) reads (24-fold genome coverage). This yielded 249 069 contigs consisting from at least two aligned reads. BLAST analysis, using the *BaSAT1* monomeric consensus sequence as the query sequence, revealed 179 contigs with a stretch of *BaSAT1* matching sequence. We excluded 31 contigs that consisted (almost) exclusively of *BaSAT1* repeats (Table S1), and additional 75 contigs, which contained also unique sequences but did not share a significant homology with *A. thaliana* genome (Table S2). The remaining 73 contigs (Table S3) mapped mainly to the euchromatic chromosome arm regions in *A. thaliana* genome (Figure 2b). There was a high concentration of the hits on the bottom arm of *A. thaliana* chromosome 2 and both arms of chromosome 5 (Figure 2b), which is consistent with the positions of HSs on chromosomes 2, 3 and 6 in

*B. antipoda* (Mandáková *et al.*, 2010). PCR-based validation of 16 *in silico*-assembled contigs confirmed 13 cases (Figure S3a). Two contigs (c134934 and c217668) mapped with their unique sequence regions to the adjacent genomic positions in the *A. thaliana* genome, suggesting that they may be separated by a single *BaSAT1* repeat array in *B. antipoda*. Indeed, PCR using primers positioned in the unique *BaSAT1* flanking sequences consistently resulted in ~7 kb a product, validating that these contigs are indeed in the same genomic location (Figure S3b). In total, nine of these contigs could be roughly placed to *B. antipoda* chromosomes based on the homology with *A. thaliana* chromosomes (Figure 2b, red arrows). To estimate the position of *BaSAT1* repeats with respect to protein coding genes, we explored the 73 *B. antipoda* BaSAT1 contigs showing homology to the *A. thaliana* genome. While, in 33 cases, sequence homology was limited to intergenic regions of *A. thaliana* genome, 40 contained sequences homologous to protein-coding genes. More detailed analysis of the latter cases revealed that 22, 14 and 4 *BaSAT1* sequence contigs were located upstream of the 5′ or downstream of the 3′ ends or directly in the coding region of a putative gene, respectively. The 22 *BaSAT1* copies found upstream of a gene were frequently located close to the translation start site (0–0.5 kb, *n* = 12; 0.5–1 kb, *n* = 5; 1–2.5 kb, *n* = 5; Table S1). Hence, *BaSAT1* satellite repeats were intermingled with single copy sequences and some occur close to, or even disrupt, protein-coding genes.

### Most *BaSAT1* repeats are DNA hypomethylated

Repeats are silenced by repressive chromatin marks in plants (Matzke and Mosher, 2014). To explore epigenetic control of *BaSAT1* repeats, we analyzed their DNA methylation and histone modifications profiles. First, we assessed the global distribution of DNA methylation by 5-methyl-2-deoxy-cytosine (5mdC)-specific immunostaining. Contrary to our expectation, there was only a weak signal over *BaSAT1* HSs on pachytene chromosomes and also in CCs of *B. antipoda* nuclei (Figure 3a; Figure S4). We excluded that the lack of signals was due to technical issues because the pericentromeric heterochromatin within the same chromosomes showed strong and continuous staining, indicating ample DNA methylation at other genome regions (Figures 3a and S4).

This prompted us to analyze *BaSAT1* DNA methylation at a single nucleotide resolution level by dideoxy-sequencing of sodium bisulfite-treated DNA. We focused on seven contigs, consisting of *BaSAT1* repeats flanked by unique sequences (Tables S2 and S3), which we were able to amplify by PCR using a combination of unique and *BaSAT1a*-specific primers. The contigs c137937 and c217668 had all cytosines in CHH context, as predicted for the *BaSAT1a* consensus sequence (Figure 1c), but other contigs also contained cytosines in symmetrical context.
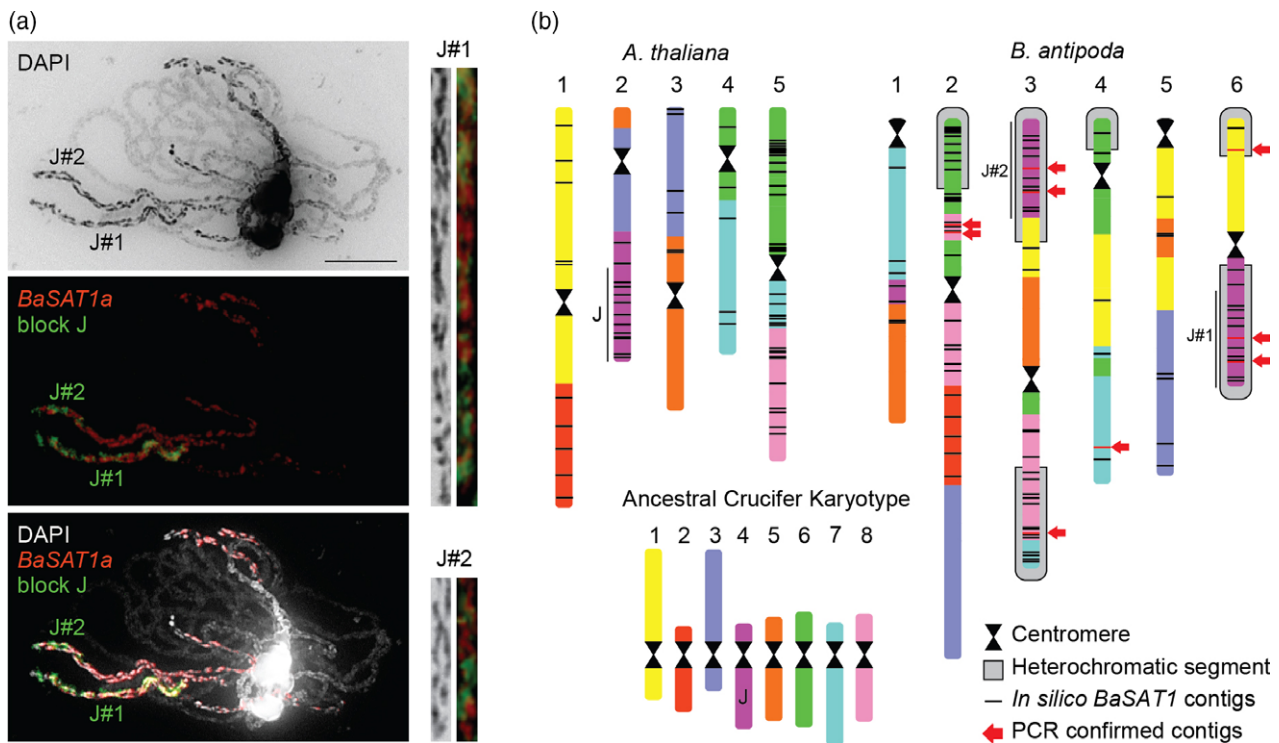
**Figure 2.** Genomic organization of heterochromatic segments (HSs) in *B. antipoda*.

(a) FISH on *B. antipoda* pachytene chromosomes using the *BaSAT1a* repeat (red) probe combined with comparative chromosome painting signal for ancient karyotype genomic block J (green). The block J appears twice due to the past polyploidization event. Chromosomes were counterstained with DAPI. The chromosomes in J#1 and J#2 were straightened using the 'straighten-curved-objects' plugin in the Image J software. Scale bar = 10 μm.

(b) Comparison of the extant karyotypes of *A. thaliana* and *B. antipoda* and the reconstructed ancestral karyotype (modified from Mandáková *et al.*, 2010). Homologous regions are indicated in the same color. Centromeres are depicted as black double-triangle structures and HSs as the gray expanded sectors below corresponding to parts of *B. antipoda* chromosomes. The genome locations of *in silico* reconstructed contigs containing *BaSAT1* are shown on *B. antipoda* and are homologous to *A. thaliana* chromosomes as black bars. Red bars indicate the position of contigs confirmed by PCR that were used for analysis of DNA methylation and histone modifications by chromatin immunoprecipitations.

The contigs c240383 and c213788 had additionally two CG sites, the contigs c118277 and c214317 had at least one cytosine in CHG context and the contig c97472 contained cytosines in all sequence contexts. Analysis of DNA methylation revealed that *BaSAT1* repeats were hypermethylated over their entire length in all contigs, except for c13721 and c217668, which were DNA hypomethylated (Figure 3b, c), suggesting that some *BaSAT1* copies may be indeed hypomethylated. We excluded this pattern to be tissue specific, as the same DNA methylation patterns were found in DNA extracted from leaves and flowers (Figure S5).

To get a representative picture of DNA methylation for more *BaSAT1* repeats, we performed DNA methylation analysis based on high-throughput bisulfite sequencing (BS-seq). The BS-seq reads were mapped to *de novo* assembled scaffolds on the *B. antipoda* genome, on which genes were predicted using Augustus software with support of *A. thaliana* TAIR10 genome annotation. This confirmed that *BaSAT1* repeats are indeed interspersed in genomic regions containing putative genes and may form complex arrays of monomers divided by spacers of variable length (Figures 4a–c and S6a). Analyzing DNA

methylation over multiple genomic regions revealed that some of the putative genes contained only CG methylation, which resembled gene body methylation (Figures S6b–c and S7), while other regions contained DNA methylation in all sequence contexts (Figure 4a–c; Figure S6a). Next we looked for DNA methylation specifically in *BaSAT1* repeats. We found 39 778 *BaSAT1* repeats on the assembled genome, out of which 7742 repeats had four or more cumulative BS-seq reads mapping to given positions; Figure 4d). Because of the absence of high quality reference genome and high repetitiveness of *BaSAT1* repeats, we estimated DNA methylation to be the percentage of methylated versus non-methylated sequenced molecules at each cytosine position covered by at least four BS-seq reads. In *BaSAT1*, there were 7.5% (out of a total 16 757) CG positions methylated, for CHG context it was 5.4% (out of a total 13 082) and for CHH context 9.8% (out of a total 39 779). For comparison, we quantified DNA methylation at (peri)centromeric regions localized *BaSAT2* repeats, which appeared DNA methylated on meiotic spreads (Figure 3a). In whole assembled genome, we found 23 594 *BaSAT2* copies, out of which 12 465 had four or more
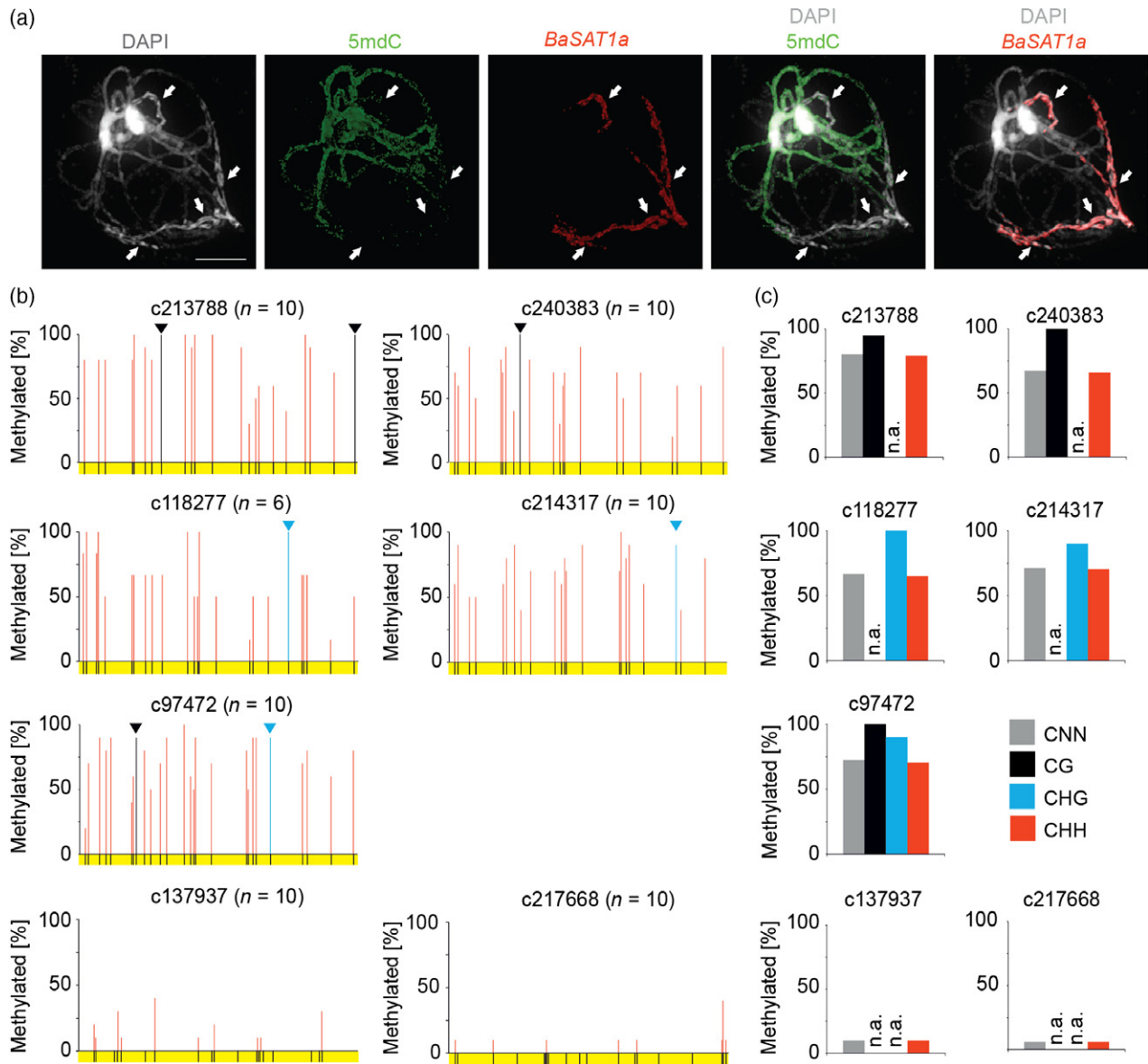
**Figure 3.** Analysis of DNA Methylation at *BaSAT1* repeats.
(a) Combination of immunostaining with 5mdC antibody (green) and FISH using a *BaSAT1a*-specific probe (red) on *B. antipoda* pachytene chromosomes. Heterochromatic segments are indicated by white arrows. Pachytenes were counterstained with DAPI. Scale bar = 10 μm.
(b) DNA methylation analysis of *BaSAT1* repeats (yellow, 174 bp) by bisulfite treatment followed by dideoxy sequencing. Each repeat was flanked by a unique sequence, into which one PCR primer was placed (see Experimental procedures). The positions of all cytosines in the reference sequence (irrespective of their methylation status) are indicated by the black bars in the yellow field. Quantitative data on the average, CG, CHG and CHH methylation are represented in gray, black, blue and red, respectively. CG and CHG methylation is further highlighted by black and blue triangles, respectively. The number of analyzed DNA molecules for each repeat is indicated as n behind the contig name.
(c) Quantitative data for (b). CNN shows the % of methylated cytosines irrespective of their sequence context; n.a., not applicable, cytosines at these contexts were absent.

cumulative BS-seq reads mapping at specific sites. For *BaSAT2*, there were 23.9% of CG, 20.0% of CHG and 27.4% of CHH methylated cytosines (out of a total 20 004, 19 345 and 23 595 sites, respectively), which is about three-fold more than for *BaSAT1* (all pairwise comparisons were P-value = 2.2E-16 in chi-squared tests; Figure 4d). For both repeats, there were no differences in frequency of DNA

methylation at different strands (Figure 4d). Hence, also BS-seq data supported DNA hypomethylation of *BaSAT1* repeats. Next, we used these data also to look into the composition of *BaSAT1* arrays with respect to both subtypes. We performed BLAST analysis using *BaSAT1a* and *BaSAT1b* consensus sequences (Figure S1) and looked whether both types occurred separately or were
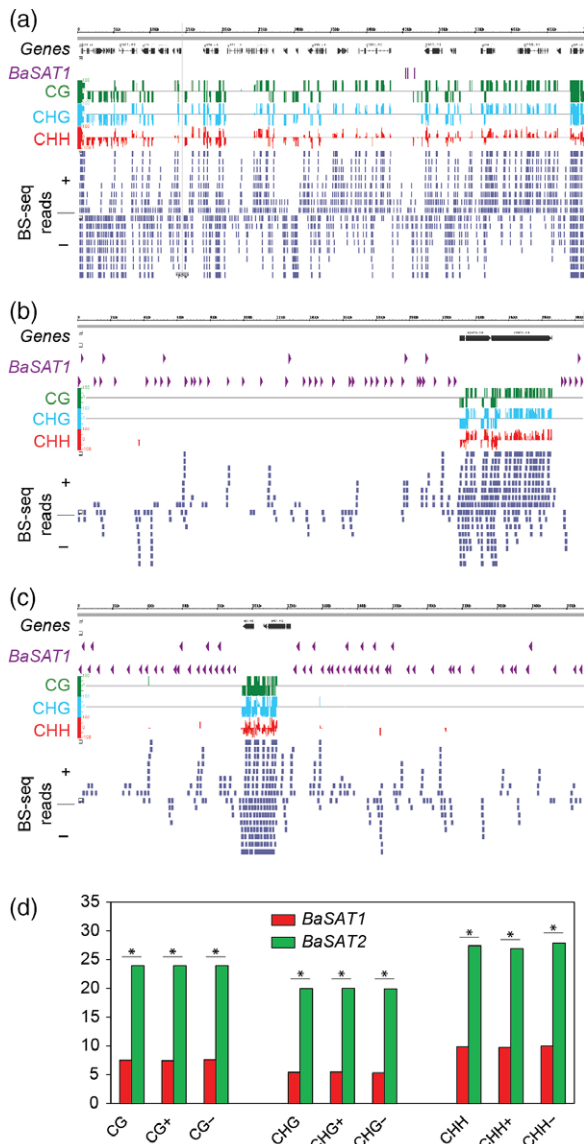
**Figure 4.** DNA methylation analysis by bisulfite sequencing (BS-seq).
(a–c) Examples of *B. antipoda* scaffolds with predicted putative genes (black), *BaSAT1* repeats (violet arrowheads) and DNA methylation information in sequence and strand-specific contexts. BS-seq reads show the coverage of the individual positions with sequencing reads on the Watson (+) and Crick (−) strands. Note that only some *BaSAT1* copies could be analyzed for DNA methylation due to limited unique mapping. (a) Shows the heavily methylated genomic region.
(b, c) Represent arrays of *BaSAT1* repeats with a single DNA methylated gene in each snapshot.
(d) Analysis of DNA methylation in *BaSAT1* and *BaSAT2* repeats based on BS-seq. We quantified the percentage of cytosine methylation in CG, CHG and CHH contexts on both and single (+ and −) DNA strands. The percentages correspond to methylated cytosine positions versus non-methylated ones. Each cytosine position had to be covered by at least four reads to be considered for analysis. All indicated comparisons were statistically significantly different (*) with a *P*-value = 2.2E-16 (chi-squared test).

intermingled. Visual inspection of multiple scaffolds revealed that, most of the time, *BaSAT1* subtypes do not intermingle (Figure S8) and only in one case we found a

*BaSAT1* array that also contained two *BaSAT1b* copies (Figure S8e).

Next, we scored for global distribution of heterochromatin- and euchromatin-specific modifications H3K9me2 and H3K4me3, respectively, in *B. antipoda* nuclei by immunostaining (Figure 5a). Intense H3K9me2 and H3K4me3 signals alternated and but a weaker H3K9me2 signal was dispersed also over the middle part of the flattened nuclei (see the overlapping images in Figure 5a). At this low resolution level, the *BaSAT1* FISH signals overlapped with both H3K4me3 and H3K9me2 signals (Figure 5a). To test this at finer scale, we determined the abundance of the H3K4me3 and H3K9me2 in specific *BaSAT1* repeats by chromatin immunoprecipitation (ChIP) along the selected contigs with known DNA methylation status (Figure 3b,c), plus contig c126293 containing a presumably euchromatic control locus *BaACTIN7* (Data S1). We found that the highly DNA methylated contigs c97472, c118277, c213788, c240383 were enriched for the H3K9me2 and depleted for the H3K4me3 modification, whereas the sparsely DNA methylated contigs c13721 and c217668 showed lower levels of H3K9me2 but high levels of H3K4me3 (Figure 5b). This suggests that individual copies of *BaSAT1* displayed either heterochromatic or euchromatic features.

### *BaSAT1*-like sequences are found in several other Australian *Microlepidieae taxa*

Unusual features of *BaSAT1* raised our curiosity about its origin and via this possibly also its dynamics in the *B. antipoda* genome. Detailed investigations into the phylogeny of the Australian *Brassicaceae* recently led to the assignment of the monotypic genus *Ballantinia* to the tribe Microlepidieae, endemic to Australasia (Heenan *et al.*, 2012). To determine whether the *BaSAT1* repeats might have originated before divergence of the Microlepidieae genera, we performed PCRs using *BaSAT1* consensus sequence-specific primers on the DNA of eight additional species (representing eight different genera) of this tribe: *Arabidella eremigena*, *Blennodia canescens*, *Cuphonotus andraeanus*, *Drabastrum alpestre*, *Harmsiodoxa puberula*, *Menkea villosula*, *Phlegmatospermum richardsii* and *Stenopetalum nutans,* as well as of *A. thaliana*. Genomic BLASTs excluded the presence of *BaSAT1*-like sequences in *A. thaliana* and therefore we used this species as control. None of the PCRs yielded a regular ladder indicative of tandem repeats, but we obtained specifically 1–1.5 kb PCR products for *P. richardsii* and an approximately 5 kb product for *M. villosula* (Figure 6a). To analyze underlying sequences, we extracted, cloned and sequenced the 1.5 kb PCR amplicon of *P. richardsii*. This revealed that (among other sequences) the band contained satellite sequences. The monomer of *PrSAT1* resembled the *BaSAT1* repeat in terms of the average length (168 bp) and A−T content
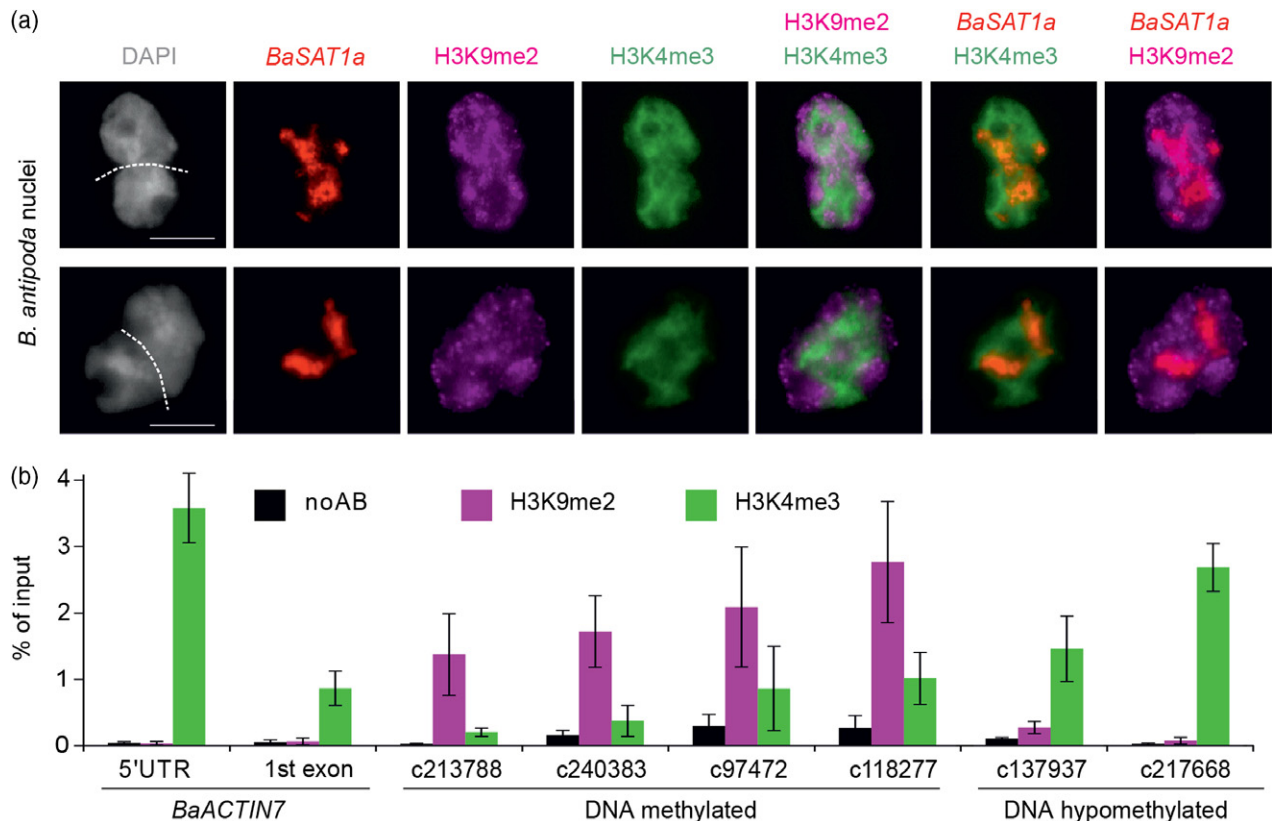
**Figure 5.** Histone modifications at *B. antipoda* heterochromatic segments.
(a) Immunostaining of *B. antipoda* nuclei with H3K9me2 (pink) and H3K4me3 antibodies (green) followed by FISH with *BaSAT1a* probe (red). Please note that both lanes showed two attached nuclei (attachment zone is indicated by the dashed line). Nuclei were counterstained with DAPI. Scale bars = 10 μm.
(b) Chromatin immunoprecipitation assaying abundance of H3K9me2 and H3K4me3 along the indicated *BaSAT1* contigs. Error bars indicate the standard deviation of two independent biological replicates. A putative *B. antipoda ACTIN 7* (*BaACTIN7*) was identified based on sequence homology to the *A. thaliana ACTIN 7* locus and used as a euchromatic control.

(79%). As observed for *BaSAT1*, a BLAST search for sequence homologs as well as the search against the PlantSAT database failed to identify repeats with a *PrSAT1* sequence similarity. Intraspecific comparison of the identified *PrSAT1* sequences revealed an average sequence similarity of 68% (ranging from 58 to 100%), which was close to the sequence variation found between *BaSAT1a* monomers (71% on average; Figures 6b and S9). FISH using the cloned *PrSAT1* sequence to *P. richardsii* mitotic chromosomes revealed one large and one small locus, suggesting that the *PrSAT1* sequences occupied specific chromosome regions in high densities, but did not span the entire chromosome arms as did *BaSAT1* repeats.

## DISCUSSION

Using a combination of low coverage genome sequencing, graph-based clustering and FISH, we identified the *BaSAT1* satellite repeat (monomer 174 bp; ca. 10% of the nuclear genome; >200 000 copies) as the principal component of the peculiar HS in the *B. antipoda* genome. Based on the survey of tandem satellite repeats in 282 species from

various kingdoms (Melters *et al.*, 2013), *BaSAT1* would be an ideal candidate for centromeric repeat sequences. However, the centromeric function of *BaSAT1* is not supported by its absence at (peri)centromeres of *B. antipoda* monocentric chromosomes, presence on both arms of chromosomes 3 and 6 (would cause dicentric chromosomes) and microscopically estimated absence on chromosomes 1 and 5 (would cause acentric chromosomes). Instead, the primary candidate for the centromeric sequence in *B. antipoda* is the second most abundant repeat *BaSAT2* with a 600-bp monomer length, which localizes to a (peri)centromeric region of all chromosomes. *BaSAT2* contains several Arabidopsis-type telomeric repeat motifs; this siituation most likely explains a strong localization of the Arabidopsis-derived telomeric repeat FISH signals within the (peri)centromeric regions of all *B. antipoda* chromosomes (Mandáková *et al.*, 2010; Majerová *et al.*, 2014).

The origin of *BaSAT1* is unclear and it is very likely to be species specific, a characteristic common to many satellite repeats (e.g. Kamm *et al.*, 1995; Ohmido *et al.*, 2000; Nouzová *et al.*, 2001). However, we found *BaSAT1*-like
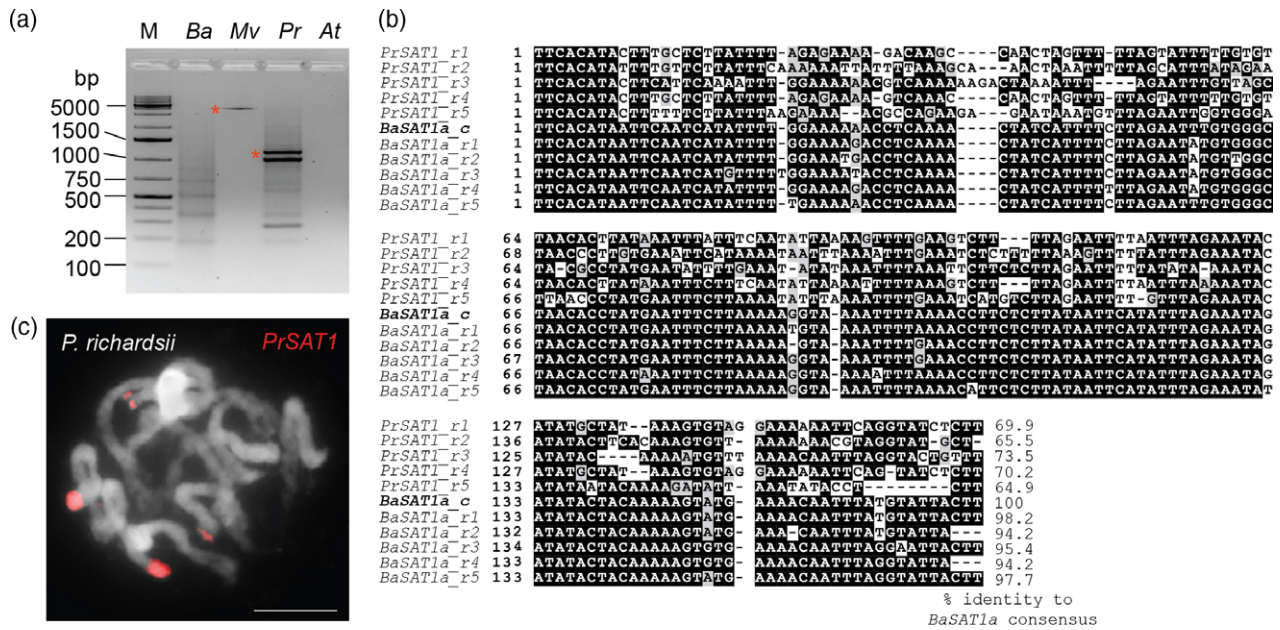
**Figure 6.** *BaSAT1*-like sequences in other Australian *Microlepidieae*.
(a) PCR using *BaSAT1a*-consensus sequence-derived primers and genomic DNA of *B. antipoda* (*Ba*), *Menkea villosula* (*Mv*), *Phlegmatospermum richardsii* (*Pr*) and *Arabidopsis thaliana* (*At*) control. Fragment sizes of the DNA marker (M) used are indicated. Asterisks indicate fragments, which were excised, cloned and sequenced.
(b) Shading of ClustalW2 alignments of five *PriSAT1* monomers of *P. richardsii* (*Pr_rep1* to *Pr_rep5*) and five *BaSAT1* monomers of *B. antipoda* (*Ba_rep1* to *Ba_rep5*) to the *BaSAT1*-consensus sequence. Shaded nucleotides were conserved in at least 50% of the aligned sequences. Identity of aligned sequences with the *BaSAT1*-consensus sequence is given after the alignment.
(c) FISH with *PrSAT1*-specific probe (red) to *P. richardsii* mitotic chromosomes counterstained with DAPI. Scale bar = 10 μm.

sequences (*PrSAT1*) occurring at two cytologically detectable genomic regions in *P. richardsii* among the eight tested species of tribe Microlepidieae. This finding suggests that *BaSAT1*-like repeats were present already in the common ancestor of at least some Australian Microlepidieae. However, at this point, we cannot exclude that other *BaSAT1*-like sequences, which are not amplified with our *BaSAT1* primers, exist in the nuclear genomes of other closely related genera.

Genomes of higher plants show a bias towards a higher A−T content, which ranges from approximately 53–67% (Barow and Meister, 2002; Lysak *et al.*, 2007; Šmarda *et al.*, 2012, 2014). The *BaSAT1* and *PrSAT1* repeats are very A−T rich (77 and 78.2%, respectively) and resemble the satellite *FriSAT1* (87% A−T) identified in the North American *Fritillaria* species (Ambrožová *et al.*, 2011). The *FriSAT1* is also present in very high copy numbers (>200 000) and can occupy large portions of the *Fritillaria* genomes, for example, up to 36% in *F. falcata*. Both *BaSAT1* and *FriSAT1* occur at many genomic positions rather than in a single or few arrays. The pattern of *BaSAT1* is even more intriguing, as it is scattered over evolutionary well conserved chromosome blocks and suggests that *BaSAT1* is capable of spreading by a currently unidentified mechanism. Speculatively, this could occur via reintegration of previously excised repeat arrays into new genomic positions and/or many microinversions, which would intermingle *BaSAT1*

with gene-rich sequences. Whether and to what extent is the amplification of *BaSAT1* repeats and related sequences allowed by the duplicated nature of *Microlepidieae* genomes (Mandáková *et al.*, 2010; Mandáková *et al.*, 2017) remains currently unknown.

Although repeats are generally fully and stably DNA methylated in plants (Mathieu *et al.*, 2003; Stroud *et al.*, 2013), recent studies from *Brassicaceae* have suggested some species-specific variability, including a reduced degree of simultaneous methylation at both cytosines in symmetrical sites in *Arabis alpina* and lacking gene body methylation in *Eutrema salsugineum* and *Conringia planisiliqua* (Willing *et al.*, 2015; Bewick *et al.*, 2016). Here, *B. antipoda* may represent yet another example. Based on the intensity of immunostaining signals, HSs appeared to be only poorly DNA methylated when compared with euchromatic chromosome arms and pericentromeric regions. A similar phenotype was described, based on cytogenetic studies, for centromeric repeats of *A. thaliana*, *Beta vulgaris*, *Zea mays* and *Oryza sativa* (Zhang *et al.*, 2008; Yan *et al.*, 2010; Zakrzewski *et al.*, 2011, 2014); however, molecular analysis by bisulfite sequencing revealed that these repeats carried a good proportion of methylated cytosines (Zakrzewski *et al.*, 2011, 2014; Schmidt *et al.*, 2014). In contrast, we found by both immunostaining and bisulfite sequencing that *BaSAT1* repeats are hypomethylated. Only about 5–10% of cytosines (depending on the

context) were methylated in *BaSAT1*, while it was about three times more (20–27%) for (peri)centromeric *BaSAT2* repeats, which appeared DNA methylated in immunostaining. The answer to which factors determine the DNA methylation status of individual *BaSAT1* repeats remains unknown. The lack of DNA methylation at most *BaSAT1* repeats is not caused by defective DNA methylation pathways, but rather by their modulation. This is suggested by the presence of dense DNA methylation in all sequence contexts at multiple genomic regions flanking *BaSAT1* repeats and also gene body methylation at many putative genes. Speculatively, DNA methylation of *BaSAT1* repeats could be influenced by the genomic neighborhood of other (DNA methylated) repeats and/or transcription over the *BaSAT1* repeat arrays, leading to the production of small interfering RNAs. However, even if existing, such small RNAs are apparently not sufficient or not abundant enough to induce genome-wide *BaSAT1* DNA methylation. Analysis of the histone modifications on six individual *BaSAT1* repeats with contrasting DNA methylation patterns revealed that DNA methylated *BaSAT1* copies are marked by repressive histone modification H3K9me2 methylation, while the low methylated ones are enriched in the permissive modification H3K4me3. Surprisingly, we also observed the H3K9me2 mark at the two repeats, which lack cytosines in CHG context. H3K9me2 is directed to specific positions by the interaction between CMT3 CHG DNA methyltransferase and SuvH4/KYP histone methyltransferase (Du *et al.*, 2012). At present it is not clear whether *B. antipoda* uses an H3K9me2 establishment mechanism independent of CHG methylation or the presence of this methylation is simply an effect of spreading from the neighboring CHG containing *BaSAT1* copies. Furthermore, our data demonstrated that individual *BaSAT1* repeats carry either heterochromatic or euchromatic features. Hence, our observations challenge the paradigm of repetitive DNA hypermethylation, and show that even highly repetitive non-coding DNA sequences can adopt euchromatic-like features in plant nuclear genomes. In conclusion, the data suggest a differential use of epigenetic pathways to control tandem repeats versus transposons.

## EXPERIMENTAL PROCEDURES

### Plant materials and growth conditions

For the origin of the analyzed species accessions, see Table S4. Seed and plant material was provided by TM and MAL. For surface sterilization, *B. antipoda* seeds were incubated in 8% sodium hypochloride solution for 10 min and subsequently washed four times in distilled water and plated on ½MS medium supplied with 15 μM gibberellic acid (GA4 + 7). Plated seeds were kept at 4°C for 48–72 h and subsequently grown under long day conditions (16 h light, 8 h dark) at 21°C. Next, 3-week-old seedlings were transferred to soil and cultivated under long-day greenhouse conditions.

### Nucleic acids isolation

DNA was prepared using the DNeasy plant mini kit (Qiagen, Hilden, Germany) or Phytopure Nucleon DNA isolation kit (GE Healthcare, Chicago, IL, USA). RNA was prepared using the RNeasy plant mini kit (Qiagen).

### Next-generation sequencing

The sequencing library of *B. antipoda* was prepared from 1 μg genomic DNA with the TruSeq DNA kit (Illumina). Library quality was assessed on a Bioanalyzer (Agilent, Santa Clara, CA, USA). The library was sequenced in a single-end 101 nt read mode using a HiSeq 2500 instrument (Illumina). The reads were quality filtered and those containing parts of the adapter sequences were discarded using FAST-X tools (http://hannonlab.cshl.edu/fastx_toolkit/).

### Identification and characterization of genomic repeats

Repeat identification by similarity-based clustering of Illumina reads was performed using local installation of the *RepeatExplorer* pipeline (Novák *et al.*, 2013), which was run on a Debian Linux server with 32 CPU cores and 64 GB RAM. In total, 1 115 000 reads were analyzed using default clustering parameter settings. The pipeline employs graphical representation of read similarities to identify clusters of frequently overlapping reads representing various repetitive elements or their parts (Novák *et al.*, 2010). In addition, it provides information about repeat quantities (estimated from the number of reads in a cluster) and outputs from BLASTn and BLASTx (Altschul *et al.*, 1990) similarity searches of our custom databases of repetitive elements and repeat-encoded conserved protein domains that aid in repeat annotation (Novák *et al.*, 2013). This information was combined and used for final manual annotation and quantification of repeats from all clusters, making up at least 0.05% of investigated genomes. Clusters containing plastid and mitochondrial sequences representing a contamination of nuclear DNA preparations by organellar DNA were excluded from the analysis, leaving 864 771 reads. Potential satellite repeats were identified based on the circular shapes of their cluster graphs (Novák *et al.*, 2010) and further analyzed using TAREAN tool of the RepeatExplorer that uses k-mer analysis of unassembled reads to reconstruct consensus sequences of tandem repeats (Novák *et al.*, 2017).

### *De novo* assembly of *B. antipoda* scaffolds and contigs

For analysis of *BaSAT1* repeat distribution on *B. antipoda* chromosomes and for local bisulfite sequencing, we performed *de novo* contig assembly using trimmed single-end reads with CLC Genomics Workbench Software (Version 5.5), using the following parameters: word size: automatic, bubble size: automatic, minimum contig length: 200. The reads were mapped back to the contigs and mismatch, insertion and deletions were penalized with 2, 3 and 3, respectively. The length fraction was set to 0.5 and similarity fraction to 0.8.

For the whole genome BS-seq analysis, we performed an additional DNA-seq experiment. Here, 500 ng of *B. antipoda* genomic DNA were dissolved in 130 μL of EB buffer and fragmented to an average size of ca. 600 bp with the S2 focused ultrasonicator (Covaris Ltd, Brighton, UK) set to the following parameters: Intensity: 3, Duty Cycle: 5%, Cycles per Burst: 200, Treatment time: 70 sec. Subsequently, fragmented DNA was concentrated using Ampure XP magnetic beads and DNA-seq libraries were constructed using the NEBNext Ultra 2 DNA library prep kit (NEB, Cat. No. E7645S)

according to the manufacturer's instructions. From these libraries, 47 345 811 PE read of 250-bp length were obtained. We assembled scaffolds and contigs using the SOAPdenovo2 program, Version 2.04 (Luo *et al.*, 2012). We used filtered paired-end and single-end DNA-seq reads with k-mer size 101 and default parameters. In total, 2 293 915 scaffolds and contigs were assembled from all the reads. The scaffolds and contigs containing *BaSAT1a* and *BaSAT1b* (Figure S1a) repeats were identified using global alignment. First, the aligned sequence files were used to generate a motif matrix file of 174 bp through the MEME application of MEME Suite (Bailey *et al.*, 2009). Then the matrix file was used to scan for repeat locations throughout the 2 293 915 assembled *B. antipoda* scaffolds and contigs using another MEME Suite application FIMO (Grant *et al.*, 2011). This yielded 35 791 scaffolds and contigs with one or more (in total 84 587) *BaSAT1* repeat regions with $q_{val} \leq 10^{-4}$.

### Chromosome preparation

Inflorescences of the analyzed accessions were fixed in ethanol: acetic acid (3:1) overnight and stored in 70% ethanol at −20°C. Selected inflorescences were rinsed in distilled water and in citrate buffer (10 mM sodium citrate, pH 4.8; 2 × 5 min) and incubated in an enzyme mix (0.3% cellulase, cytohelicase, and pectolyase; all Sigma-Aldrich) in citrate buffer at 37°C for 3–6 h. Individual flower buds were disintegrated on a microscope slide in a drop of citrate buffer and 15–30 μL of 60% acetic acid. The suspension was spread on a hot plate at 50°C for 0.5–2 min. Chromosomes were fixed by adding 100 μL of ethanol:acetic acid (3:1). The slide was dried with a hair dryer, post-fixed in 4% formaldehyde dissolved in distilled water for 10 min, and air dried. Chromosome preparations were treated with 100 μg/mL RNase in 2× sodium saline citrate (SSC; 20× SSC: 3 M sodium chloride, 300 mM trisodium citrate, pH 7.0) for 60 min and with 0.1 mg/mL pepsin in 0.01 M HCl at 37°C for 5 min; then post-fixed in 4% formaldehyde in 2× SSC, and dehydrated in an ethanol series (70, 90, and 100%, 2 min each).

### Fluorescence *in situ* hybridization

Satellite repeats of *Ballantinia* (*BaSAT1* and *BaSAT2*) and *Phlegmatospermum* (*PrSAT1*), and *Arabidopsis thaliana* BAC clones corresponding to genomic block J of the Ancestral Crucifer Karyotype (ACK; Schranz *et al.*, 2006; Lysak *et al.*, 2016) were used as FISH probes. All DNA probes were labeled with biotin-dUTP or digoxigenin-dUTP by nick translation as described (Mandáková and Lysak, 2016). Selected labeled DNA probes were pooled together, ethanol precipitated, dissolved in 20 μL of 50% formamide, 10% dextran sulfate in 2× SSC and pipetted onto microscopic slides. The slides were heated at 80°C for 2 min and incubated at 37°C overnight. Post-hybridization washing was performed in 20% formamide in 2× SSC at 42°C. Hybridized probes were visualized through fluorescently labeled antibodies against biotin-dUTP and digoxigenin-dUTP (Mandáková and Lysak, 2016). Chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI, 2 μg/mL) in Vectashield antifade. Fluorescence signals were analyzed and photographed using a Zeiss Axioimager epifluorescence microscope and a CoolCube camera (MetaSystems, Heidelberg, Germany). Individual images were merged and processed using Photoshop CS software (Adobe Systems, San Jose, CA, USA). Pachytene chromosomes in Figure 2 were straightened using the 'straighten-curved-objects' plugin in the Image J software (Kocsis *et al.*, 1991).

### 5-Methyl-2′-deoxy-cytosine (5mdC) immunodetection

For immunostaining of 5mdC, standard chromosome preparations (see above) were used. A denaturation mixture containing 20 μL of 50% formamide, 10% dextran sulfate in 2× SSC was pipetted onto microscopic slides. The slides were heated at 80°C for 2 min and washed in 2× SSC (2 × 5 min). Slides were blocked for 30 min with 5% BSA solution (5% bovine serum albumin, 0.2% Tween-20 in 4× SSC) at 37°C for 30 min and then incubated with 100 μL of primarily antibody against 5mC (diluted 1:100, Diagenote) at 37°C for 30 min. After washing two times in 2× SSC the primary antibody was detected with the secondary antibody coupled with AlexaFluor488 (diluted 1:200, Invitrogen) at 37°C for 30 min followed by washing two times in 2× SSC and a dehydration in an ethanol series (70, 90, and 100%, 2 min each). Chromosomes were counterstained with DAPI, fluorescence signals analyzed and photographed, and slides rehybridized by satellite probes as described above.

### Histone immunolabeling

Leaf tissue (0.5–1 g) with 0.5 mL of NIB buffer (10 mM Tris−HCl, 10 mM EDTA, 100 mM KCl, 0.5 M sucrose, 4 mM spermidine, 1 mM spermine, 0.1% 2-mercaptoethanol) was placed into a Petri dish on ice and chopped to a fine suspension with the razor blade. The suspension was pipetted into an Eppendorf tube, fixed in an equal volume of 4% paraformaldehyde on ice for 20 min, filtered through 50 and 20 μm mesh filters and centrifuged at 595 *g* at 4°C for 3 min. The supernatant was removed and the pellet with nuclei resuspended in 40 μL of NIB. Then, 2 μL of the suspension were pipetted onto a slide, dried at 4°C for 1 h and post-fixed in 4% formaldehyde in 2× SSC for 30 min. Slides were blocked for 30 min with 5% BSA solution at 37°C and then incubated with 100 μL of primarily antibodies against H3K4met3 and H3K9met2 (diluted 1:100, Abcam, Cambridge, UK) at 37°C for 2 h. After washing two times in 2× SSC the primary antibodies were detected with the secondary antibodies coupled with AlexaFluor488 (diluted 1:200, Invitrogen) and Cy5 (diluted 1:100, Jackson ImmunoResearch) at 37°C for 30 min followed by washing two times in 2× SSC. Chromosomes were counterstained with DAPI, fluorescence signals analyzed and photographed, and slides re-hybridized by satellite probes as described above.

### DNA methylation analysis

For local DNA methylation analysis, 150–200 ng of *B. antipoda* genomic DNA was treated with sodium bisulfite using the Epitect Bisulfite Kit (Qiagen). PCR fragments were amplified for 32–35 cycles using MethylTaq DNA polymerase (Diagenode, Seraing, Belgium) according to manufacturers' recommendations, purified with QIAQuick PCR purification kit (Qiagen) and cloned into the pJet1.2 vector using the ClonJet PCR cloning kit (Thermo Scientific, Waltham, MA, USA). Colony PCR was performed to identify positive clones and the positive plasmids were isolated using the NucleoSpin Plasmid Mini Prep Kit (Macherey Nagel, Düren, Germany) and sequenced on an 3730XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The sequences were trimmed, aligned with the ClustalW algorithm and analyzed using CyMATE (Hetzl *et al.*, 2007).

For genome-wide DNA methylation analysis, 1000 ng of *B. antipoda* genomic DNA was isolated and provided to the Max Planck Genome Centre in Cologne, Germany (https://mpgc.mpipz.mpg.de/home/) for library construction. The genomic DNA was sheared to fragments of ca. 400 bp using a S2 focused

ultrasonicator (Covaris Inc., Brigthon). A BS-seq library was constructed using the Bioo Scientific NEXTFLEX® Bisulfite Library Prep Kit (Perkin Elmer, Waltham, MA, USA) according to manufacturer's recommendations. The library was sequenced as 150-bp long paired-end reads. The reads were mapped to 35 791 *BaSAT1* repeats containing scaffolds and contigs using Bismark aligner software (Krueger and Andrews, 2011). Then, we used the Bismark methylation extractor (Krueger and Andrews, 2011) for strand-specific identification of methylated cytosines. This software yielded a bedgraph file of 5mdC, in which each methylation was reported in terms of location, context, and frequency. The scaffolds contained in total 84 587 *BaSAT1* repeats (i.e. often multiple repeats per one scaffold). For DNA methylation analysis, we identified cytosines in BaSAT1 and BaSAT2 repeats covered by at least four BS-seq reads, which resulted in 7742 and 12 463 analyzable *BaSAT1* and BaSAT2 copies, respectively. The percentage of methylated and non-methylated positions was calculated for each cytosine and then summed up for the whole repeat.

### ChIP

ChIP was done as described (Gendrel *et al.*, 2005) with modifications: 3 g of leaves of 12-week-old soil grown plants were harvested. Crosslinking was performed in 37 mL of 1% (w/v) formaldehyde solution under vacuum for 20 min and subsequently quenched with 2.5 mL of 2 M glycine solution (final concentration 0.125 M) under vacuum for 7 min. Crosslinked material was snap frozen in liquid nitrogen, homogenized under liquid nitrogen, suspended in 30 mL, filtered through four layers of Miracloth and subsequently centrifuged for at 2000 $g$ for 20 min at 4°C. After resuspension of the pellet in 1 mL of extraction buffer 2 the solution was centrifuged at 17 000 $g$ for 15 min at 4°C. The resulting pellet was resuspended in 300 μL of extraction buffer 3, overlayed on 300 μL of extraction buffer 3 and centrifuged at 17 000 $g$ for 1 h at 4°C. This step was repeated once. The nuclei pellet was suspended in 300 μL of ice-cold nuclei lysis buffer and chromatin was sheared at 4°C using a Diagenode Disruptor for six cycles with 30 sec of high energy sonication and a 30 sec break. Subsequently a centrifugation at 17 000 $g$ for 10 min at 4°C was performed to remove nuclear debris. An aliquot of the chromatin extract was set aside to serve as the input control. The remaining extract was diluted 1:10 in ChIP dilution buffer. The chromatin solution was divided into four aliquots, 40 μL of Protein A Magnetic Sepharose beads (GE Healthcare) per mL were added and incubated for 45 min at 4°C with slight agitation. Subsequently the solution was centrifuged at 12 000 $g$ for 30 sec at 4°C and the supernatant was transferred to a fresh tube. Three microlitres of the following antibodies were added were added to the respective tubes. H3K4me3: #39159 Histone H3 trimethyl Lys4 Rabbit pAB, Activemotif, H3K9me2: #720092 dimethyl-histone H3 Lys9 pAB, (Invitrogen). One aliquot served as the no Ab control. The immunoprecipitation reaction was incubated overnight at 4°C under slight agitation. After incubation, 40 μL/mL of Protein A Magnetic Agarose beads were added the solution, incubated for 1 h and centrifuged for 10 min at 12 000 $g$ to pellet the beads. Beads were washed twice 5 min each with low or high salt buffer (150 and 500 mM NaCL, respectively; plus 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris−HCl (pH 8.1)); LiCl wash buffer: 0.25 LiCl, 1% NP40, 1% Na deoxycholate, 1 mM EDTA, 10 mM Tris−HCl (pH 8.1) and TE buffer: 10 mM Tris−HCl (pH 8.0), 1 mM EDTA. The DNA was eluted twice by incubation with 250 μL elution buffer (Qiagen) at 65°C for 15 min.

Quantitative (real-time) PCR was performed using the QPCR Green Master Mix Fluorescein Kit (BiotechRabbit, BR0501203, Berlin, Germany) in 12 μL QPCR reaction according to

manufacturer's protocols. The samples were amplified using a CFX384 Touch real-time PCR Detection System (Bio-Rad, Hercules, CA, USA), and quantified with a calibration line made with DNA isolated from crosslinked, sonicated chromatin. With all experiments, no-template controls, No Ab controls and input samples were taken along for every primer set used. As the control, abundance of the respective histone modifications at the 5′UTR and the first exon of a putative *B. antipoda Actin7* (*BaActin7*) gene was assayed. The *BaActin7* sequence was identified by BLAST analysis of the *B. antipoda* contig library using the *A. thaliana* Actin7 gene as query.

### Primers

All primers and oligonucleotides used in this study are defined in Table S5.

### Sequence deposition

Adapter-trimmed raw sequencing reads generated in this study are deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under the accession number PRJEB21350 (for the content see Table S6).

### CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

### AUTHOR CONTRIBUTIONS

GTHV prepared samples for genome sequencing. AF performed all molecular and TM all cytogenetic experiments. PN and JM performed graph-based clustering analysis and generated repeat consensus sequences. KN performed genome assembly, *in silico* repeat analysis and analyzed bisulfite sequencing data. AP, AF, MAL and JM designed the experiments and wrote the manuscript.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Consensus sequences of the *BaSAT1a* and *BaSAT1b* satellite repeats.

**Figure S2.** Consensus sequence of the *BaSAT2* repeat.

**Figure S3.** PCR confirmation of *BaSAT1* containing contigs.

**Figure S4.** Immunostaining reveals low DNA methylation of *BaSAT1* repeats.

**Figure S5.** *BaSAT1* bisulfite sequencing analysis at different developmental stages.

**Figure S6.** Examples of DNA methylation in the *B. antipoda* genome.

**Figure S7.** Additional examples of DNA methylation in the *B. antipoda* genome.

**Figure S8.** *BaSAT1a* and *BaSAT1b* repeats are organized into separate arrays.

**Figure S9.** *PrSAT1* sequence analysis.

**Table S1.** *BaSAT1* repeat only contigs.

**Table S2.** Unchanchored *BaSAT1* contigs.

**Table S3.** Homeology of *BaSAT1* contigs to the *A. thaliana* genome.

**Table S4.** Origin of materials from Australian Brassicaceae species used in this study.

**Table S5.** Oligonucleotides used in this study.

**Table S6.** Content of the dataset PRJEB21350 deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena).

**Data S1.** Sequences of assembled contigs (fasta-format).

# REFERENCES

**Ali, H.B.M., Lysak, M.A. and Schubert, I.** (2005) Chromosomal localization of rDNA in the Brassicaceae. *Genome*, **48**, 341–346.

**Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D.** (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

**Ambrožová, K., Mandáková, T., Bureš, P., Neumann, P., Leitch, I.J., Koblížková, A., Macas, J. and Lysak, M.A.** (2011) Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of Fritillaria lilies. *Ann. Bot.* **107**, 255–268.

**Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S.** (2009) MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208.

**Barow, M. and Meister, A.** (2002) Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. *Cytometry*, **47**, 1–7.

**Baubec, T., Finke, A., Mittelsten Scheid, O. and Pecinka, A.** (2014) Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Rep.* **15**, 446–452.

**Belele, C.L., Sidorenko, L., Stam, M., Bader, R., Arteaga-Vazquez, M.A. and Chandler, V.L.** (2013) Specific tandem repeats are sufficient for paramutation-induced trans-generational silencing. *PLoS Genet.* **9**, e1003773.

**Bewick, A.J., Ji, L., Niederhuth, C.E.** *et al.* (2016) On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl Acad. Sci. USA*, **113**, 9111–9116.

**Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M. and Mittelsten Scheid, O.** (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet.* **10**, e1004115.

**Cowan, C.R., Carlton, P.M. and Cande, W.Z.** (2001) The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol.* **125**, 532–538.

**Dawe, R.K., Lowry, E.G., Gent, J.I.** *et al.* (2018) A Kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. *Cell*, **173**, 839–850.e18.

**Devos, K.M., Brown, J.K.M. and Bennetzen, J.L.** (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res.* **12**, 1075–1079.

**Du, J., Zhong, X., Bernatavichute, Y.V.** *et al.* (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167–180.

**Fransz, P., de Jong, J.H., Lysak, M., Castiglione, M.R. and Schubert, I.** (2002) Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl Acad. Sci. USA*, **99**, 14584–14589.

**Fransz, P., Soppe, W. and Schubert, I.** (2003) Heterochromatin in interphase nuclei of Arabidopsis thaliana. *Chromosome Res.* **11**, 227–240.

**Fransz, P., ten Hoopen, R. and Tessadori, F.** (2006) Composition and formation of heterochromatin in Arabidopsis thaliana. *Chromosome Res.* **14**, 71–82.

**Fransz, P., Linc, G., Lee, C.R.** *et al.* (2016) Molecular, genetic and evolutionary analysis of a paracentric inversion in Arabidopsis thaliana. *Plant J.* **88**, 159–178.

**Fu, F.-F., Dawe, R.K. and Gent, J.I.** (2018) Loss of RNA-directed DNA methylation in maize chromomethylase and DDM1-type nucleosome remodeler mutants. *Plant Cell*, **30**, 1617–1627.

**Garrido-Ramos, M.A.** (2015) Satellite DNA in plants: more than just rubbish. *Cytogenet Genome Res.* **146**, 153–170.

**Gendrel, A.-V., Lippman, Z., Martienssen, R. and Colot, V.** (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat. Methods*, **2**, 213–218.

**Gent, J.I., Madzima, T.F., Bader, R., Kent, M.R., Zhang, X., Stam, M., McGinnis, K.M. and Dawe, R.K.** (2014) Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. *Plant Cell*, **26**: 4903, LP-4917.

**Grant, C.E., Bailey, T.L. and Noble, W.S.** (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.

**Gregory, T.R.** (2005) The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann. Bot.* **95**, 133–146.

**Hawkins, J.S., Proulx, S.R., Rapp, R.A. and Wendel, J.F.** (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc. Natl Acad. Sci. USA*, **106**, 17811–17816.

**Hay, A.S., Pieper, B., Cooke, E.** *et al.* (2014) Cardamine hirsuta: a versatile genetic system for comparative studies. *Plant J.* **78**, 1–15.

**Heenan, P.B., Goeke, D.F., Houliston, G.J. and Lysak, M.A.** (2012) Phylogenetic analyses of ITS and rbcL DNA sequences for sixteen genera of Australian and New Zealand Brassicaceae result in the expansion of the tribe Microlepidieae. *Taxon*, **61**, 970–979.

**Heslop-Harrison, J.S. and Schwarzacher, T.** (2011) Organisation of the plant genome in chromosomes. *Plant J.* **66**, 18–33.

**Hetzl, J., Foerster, A.M., Raidl, G. and Mittelsten Scheid, O.** (2007) CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51**.

**Hu, T.T., Pattyn, P., Bakker, E.G.** *et al.* (2011) The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481.

**Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G.** *et al.* (2013) Architecture and evolution of a minute plant genome. *Nature*, **498**, 94–98.

**Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. and Paszkowski, J.** (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, **472**, 115–119.

**Kamm, A., Galasso, I., Schmidt, T. and Heslop-Harrison, J.S.** (1995) Analysis of a repetitive DNA family from Arabidopsis arenosa and relationships between Arabidopsis species. *Plant Mol. Biol.* **27**, 853–862.

**Kocsis, E., Trus, B.L., Steer, C.J., Bisher, M.E. and Steven, A.C.** (1991) Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment. *J. Struct. Biol.* **107**, 6–14.

**Krueger, F. and Andrews, S.R.** (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.

**Lisch, D.** (2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49–61.

**Luo, R., Liu, B., Xie, Y.** *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, **1**, 18.

**Lysak, M.A., Koch, M.A., Pecinka, A. and Schubert, I.** (2005) Chromosome triplication found across the tribe Brassiceae. *Genome Res.* **15**, 516–525.

**Lysak, M.A., Cheung, K., Kitschke, M. and Bureš, P.** (2007) Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiol.* **145**, 402–410.

**Lysak, M.A., Mandáková, T. and Schranz, M.E.** (2016) Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Curr. Opin. Plant Biol.* **30**, 108–115.

**Macas, J., Navrátilová, A. and Mészáros, T.** (2003) Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. *Chromosoma*, **112**, 152–158.

**Macas, J., Koblížková, A., Navrátilová, A. and Neumann, P.** (2009) Hypervariable 3′ UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. *Gene*, **448**, 198–206.

**Majerová, E., Mandáková, T., Vu, G.T.H., Fajkus, J., Lysak, M.A. and Fojtová, M.** (2014) Chromatin features of plant telomeric sequences at terminal vs. internal positions. *Front. Plant Sci.* **5**, 593.

**Mandáková, T. and Lysak, M.A.** (2008) Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (Brassicaceae). *Plant Cell*, **20**, 2559–2570.

**Mandáková, T. and Lysak, M.A.** (2016). Chromosome preparation for cytogenetic analyses in Arabidopsis. In *Curr Protocols Plant Biol.* John Wiley & Sons, Inc., pp. 43–51.

**Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K. and Lysak, M.A.** (2010) Fast diploidization in close mesopolyploid relatives of Arabidopsis. *Plant Cell*, **22**, 2277–2290.

**Mandáková, T., Pouch, M., Harmanová, K., Zhan, S.H., Mayrose, I. and Lysak, M.A.** (2017) Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Mol. Ecol.* **26**, 6445–6462.

**Mari-Ordonez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V. and Voinnet, O.** (2013) Reconstructing de novo silencing of an active plant retrotransposon. *Nat. Genet.* **45**, 1029–1039.

**Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.-V., Colot, V., Schubert, I. and Tourmente, S.** (2003) Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in Arabidopsis. *Plant Cell*, **15**, 2929–2939.

**Matzke, M.A. and Mosher, R.A.** (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408.

**Mehrotra, S. and Goyal, V.** (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinformatics*, **12**, 164–171.

**Melters, D.P., Bradnam, K.R., Young, H.A.** *et al.* (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10–R10.

**Nouzová, M., Neumann, P., Navrátilová, A., Galbraith, D.W. and Macas, J.** (2001) Microarray-based survey of repetitive genomic sequences in *Vicia* spp. *Plant Mol. Biol.* **45**, 229–244.

**Novák, P., Neumann, P. and Macas, J.** (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, **11**, 378.

**Novák, P., Neumann, P., Pech, J., Steinhaisl, J. and Macas, J.** (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, **29**, 792–793.

**Novák, P., Robledillo, L.A., Koblížková, A., Vrbová, I., Neumann, P. and Macas, J.** (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **45**, e111.

**Ohmido, N., Kijima, K., Akiyama, Y., de Jong, J.H. and Fukui, K.** (2000) Quantification of total genomic DNA and selected repetitive sequences reveals concurrent changes in different DNA families in indica and japonica rice. *Mol. Gen. Genet.* **263**, 388–394.

**Piegu, B., Guyot, R., Picault, N.** *et al.* (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. *Genome Res.* **16**, 1262–1269.

**Pietzenuk, B., Markus, C., Gaubert, H., Bagwan, N., Merotto, A., Bucher, E. and Pecinka, A.** (2016) Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. *Genome Biol.* **17**, 209.

**Plohl, M., Luchetti, A., Meštrović, N. and Mantovani, B.** (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene*, **409**, 72–82.

**Schmidt, M., Hense, S., Minoche, A.E., Dohm, J.C., Himmelbauer, H., Schmidt, T. and Zakrzewski, F.** (2014) Cytosine methylation of an ancient satellite family in the wild beet Beta procumbens. *Cytogenet Genome Res.* **143**, 157–167.

**Schnable, P.S., Ware, D. and Fulton, R.S.** (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

**Schranz, M.E., Lysak, M.A. and Mitchell-Olds, T.** (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* **11**, 535–542.

**Seymour, D.K., Koenig, D., Hagmann, J., Becker, C. and Weigel, D.** (2014) Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. *PLoS Genet.* **10**, e1004785.

**Šmarda, P., Bureš, P., Šmerda, J. and Horová, L.** (2012) Measurements of genomic GC content in plant genomes with flow cytometry: a test for reliability. *New Phytol.* **193**, 513–521.

**Šmarda, P., Bureš, P., Horová, L., Leitch, I.J., Mucina, L., Pacini, E., Tichý, L., Grulich, V. and Rotreklová, O.** (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc. Natl Acad. Sci. USA*, **111**, E4096–E4102.

**Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V. and Jacobsen, S.E.** (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*, **152**, 352–364.

**Tiang, C.-L., He, Y. and Pawlowski, W.P.** (2012) Chromosome organization and dynamics during interphase, mitosis, and meiosis in plants. *Plant Physiol.* **158**, 26–34.

**Vu, G.T.H., Schmutzer, T., Bull, F.** (2015) Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. *Plant Genome* **8**. https://doi.org/10.3835/plantgenome2015.04.0021.

**Willing, E.M., Rawat, V., Mandáková, T.** *et al.* (2015) Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. *Nat. Plants*, **1**, 1–5.

**Willing, E.-M., Piofczyk, T., Albert, A., Winkler, B.J., Schneeberger, K. and Pecinka, A.** (2016) UVR2 ensures trans-generational genome stability under simulated natural UV-B in Arabidopsis thaliana. *Nat. Commun.* **7**, 13522.

**Yadav, R.K., Girke, T., Pasala, S., Xie, M. and Reddy, G.V.** (2009) Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. *Proc. Natl Acad. Sci. USA*, **106**, 4941–4946.

**Yan, H., Kikuchi, S., Neumann, P., Zhang, W., Wu, Y., Chen, F. and Jiang, J.** (2010) Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *Plant J.* **63**, 353–365.

**Zakrzewski, F., Weisshaar, B., Fuchs, J., Bannack, E., Minoche, A.E., Dohm, J.C., Himmelbauer, H. and Schmidt, T.** (2011) Epigenetic profiling of heterochromatic satellite DNA. *Chromosoma*, **120**, 409–422.

**Zakrzewski, F., Schubert, V., Viehoever, P., Minoche, A.E., Dohm, J.C., Himmelbauer, H., Weisshaar, B. and Schmidt, T.** (2014) The CHH motif in sugar beet satellite DNA: a modulator for cytosine methylation. *Plant J.* **78**, 937–950.

**Zhang, W., Lee, H.-R., Koo, D.-H. and Jiang, J.** (2008) Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell*, **20**, 25–34.

# Publikace 5

# ARTICLE

# UVR2 ensures transgenerational genome stability under simulated natural UV-B in *Arabidopsis thaliana*

Eva-Maria Willing[1,*], Thomas Piofczyk[2,*], Andreas Albert[3], J. Barbro Winkler[3], Korbinian Schneeberger[1] & Ales Pecinka[2]

Ground levels of solar UV-B radiation induce DNA damage. Sessile phototrophic organisms such as vascular plants are recurrently exposed to sunlight and require UV-B photoreception, flavonols shielding, direct reversal of pyrimidine dimers and nucleotide excision repair for resistance against UV-B radiation. However, the frequency of UV-B-induced mutations is unknown in plants. Here we quantify the amount and types of mutations in the offspring of *Arabidopsis thaliana* wild-type and UV-B-hypersensitive mutants exposed to simulated natural UV-B over their entire life cycle. We show that reversal of pyrimidine dimers by UVR2 photolyase is the major mechanism required for sustaining plant genome stability across generations under UV-B. In addition to widespread somatic expression, germline-specific UVR2 activity occurs during late flower development, and is important for ensuring low mutation rates in male and female cell lineages. This allows plants to maintain genome integrity in the germline despite exposure to UV-B.

[1] Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Cologne, Germany. [2] Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Cologne, Germany. [3] Research Unit Environmental Simulation, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to K.S. (email: schneeberger@mpipz.mpg.de) or to A.P. (email: pecinka@mpipz.mpg.de).

Plants require sunlight for photosynthesis and developmental regulation[1]. However, ground levels of solar radiation also contain a low proportion of UV-B radiation (UV-B, 280–315 nm), which has multiple effects on plants including photomorphogenic and damaging responses[2–4]. Photomorphogenic responses are triggered upon UV-B perception by UV-B-RESISTANCE 8 (UVR8)[2,5]. UV-B-irradiated UVR8 homodimers will monomerize and bind COP1 E3 ubiquitin ligase. Reduced COP1 activity will allow accumulation of HY5 transcription factor and will trigger UV-B transcriptional response of $\sim 100$ target genes and more compact plant growth, including, e.g., reduced plant height and shorter petioles[4]. Furthermore, low UV-B levels boost accumulation of flavonoid pigments, in a TRANSPARENT TESTA 4 (TT4)-dependent manner, which will build up a protective sunscreen layer contributing to UV-B acclimation and even protection against other stresses[5,6]. Higher natural and, in particular, laboratory-applied UV-B doses cause damage[3,7,8]. This involves a burst of reactive oxygen species, damages to cell membranes, proteins and DNA. The major types of UV-B-induced DNA damage are pyrimidine dimers and, to a lower extent, also DNA strand breaks[9–11]. Pyrimidine dimers are non-native bonds between two pyrimidines (cytosine and thymine). They disturb DNA structure, interfere with replication and transcription, and are therefore generally repaired[12]. The cyclobutane pyrimidine dimers (CPDs; 75–90% of all pyrimidine dimers) and 6,4 pyrimidine-pyrimidones ((6-4)PPs; 10–25% of all pyrimidine dimers) are directly reverted by UV-B-RESISTANCE 2 (UVR2) and UV-B-RESISTANCE 3 (UVR3) photolyases, respectively, in somatic tissues[13,14]. An alternative repair pathway common to all eukaryotes involves nucleotide excision repair (NER). In A. thaliana, loss of NER-associated endonuclease UV-B HYPERSENSITIVE 1 (UVH1), an orthologue of human XERODERMA PIGMENTOSUM COMPLEMENTATION GROUP F (XPF), leads to failures in repair of UV-B-induced lesions and reduced growth in response to UV-B treatment[15,16]. Owing to the low UV-B penetration into plant tissues through flavonoid layer[17], most of the UV-B-induced mutations are to be expected in the epidermal cells. However, there is some evidence that UV-B may penetrate also into deeper meristematic cell layers as even low UV-B increases genome instability in the plant germline[11]; however, the precise frequencies of UV-B-induced mutations and their molecular spectra remain unknown in plants.

Here we determined mutation frequencies in germline DNA of A. thaliana wild-type and UV-B-hypersensitive mutants exposed to UV-B treatment by a combination of whole-genome sequencing and genetic analyses. We found that mutations induced by the UV-B treatment have specific spectra, preferentially occur in particular sequence contexts and have other characteristics that differentiate them from spontaneous mutations. Furthermore, we show that direct reversal by UVR2 photolyase is the key pathway limiting the frequency of UV-B treatment-induced mutations in the DNA of germline cells. We localized this repair activity into late flower development after the split of male- and female-specific cell lineages.

## Results

### Effects of simulated solar UV-B on A. thaliana growth. Wild-type plants and six mutant genotypes uvr8, tt4, uvh1, uvr2, uvr3 and uvr2 uvr3 found as UV-B- and/or UV-C-hypersensitive in previous studies[6,15,18,19] were cultivated during their entire life cycle in sun simulators[20] for up to three generations without UV-B (hereafter as 'control') and with a biologically effective UV-B radiation (UV-B$_{BE}$) normalized at 300 nm (ref. 21) of 100, 150 and 300 mW m$^{-2}$ (Fig. 1a and Supplementary Fig. 1a–c). Owing

to the filtering conditions used, this UV-B treatment did lead to more UV-A than in the control treatment. However, the amount of UV-A radiation in the control treatment reached up to 80% and more for wavelengths greater than 360 nm compared to the UV-B treatments. Below 360 nm the transmission decreased due to the transmission characteristics of the filter glass, therefore, the UV-A radiation is reduced to about 10% at 330 nm compared to the UV-B treatments. The UV-B treatments resembled natural conditions during the main A. thaliana-growing season (April/May) along the European north-south UV-B cline at 60°N, 52°N and 40°N, which can be approximated to Helsinki, Berlin, and Madrid, respectively. Wild-type and all mutant genotypes showed comparable growth at rosette stages under control conditions (Fig. 1b). Under the highest simulated natural UV-B, wild-type and uvr8 plants did not show significantly reduced rosette diameter, while tt4, uvr2, uvr3, uvr2 uvr3 and uvh1 mutant plants did (t-test P values: 5.390E − 01, 9.113E − 01, 4.3E − 06, 1.6E − 16, 4.4E − 02, 2.6E − 16 and 8.2E − 03, respectively; Fig. 1b). This suggested that not all A. thaliana mutants found to be UV-B- and/or UV-C-hypersensitive in laboratory would show similar phenotypes under natural UV-B conditions.

### Frequency of mutations induced by UV-B treatment. The seeds of control and UV-B-treated plants were grown under non-UV-B conditions and whole genomes of 146 offspring plants, typically five per genotype and treatment, were sequenced (Supplementary Fig. 2 and Supplementary Data 1). This revealed a total of 2,497 novel single-base substitutions and 22 one-to-four base pair deletions. Using di-deoxy sequencing, we confirmed 58 out of 59 randomly selected mutations, suggesting a 1.7% false-positive discovery rate in our analysis (Supplementary Data 2 and Methods). A false-negative mutation discovery rate was estimated to be 0.15% by simulations (see Methods).

Wild-type plants without UV-B treatment accumulated on average 2.6, 2.0 and 2.4 spontaneous mutations per haploid genome and generation (hereafter as 'mutations') in the first (Fig. 1c), the second and the third generations (generation average 2.3), corresponding to 2.2, 1.7 and 2.0 $\times 10^{-8}$ mutations per site, respectively (Supplementary Data 1). Similar numbers of novel mutations (2.0–5.7) were observed in the progenies of control uvr8, tt4, uvr2, uvr3 and uvr2 uvr3 plants (Fig. 1c and Supplementary Data 1). In contrast, compromised NER in uvh1 plants resulted in 20.3 mutations. This represented 7.8-fold increase (Fisher's exact test, P = 4.9E − 12) compared with wild-type and illustrated importance of NER for general genome stability in A. thaliana.

Treatment with 100, 150 and 300 mW m$^{-2}$ induced 3.3, 5.0 and 2.8 mutations, respectively, per haploid genome and generation in wild-type plants (Supplementary Fig. 3a). Subsequently, the UV-B$_{BE}$ of 300 mW m$^{-2}$ was used as the standard UV-B treatment. Loss of UVR8 and TT4 functions did not significantly change the mutation rates (5.6 versus 7.8 and 5.7 versus 6.7 mutations under control and UV-B; Fisher' exact test P = 0.2203 and 0.6455, respectively; Fig. 1c). In UV-B-treated uvh1 plants, we found 27.4 new mutations, which represented a significant 1.3-fold increase compared with 20.3 new mutations under control conditions (Fisher's exact test, P = 0.03772).

The only drastic increase in mutation rate in a single mutant was observed in the progeny of UV-B-irradiated uvr2 plants containing on average 64.3 new mutations (Fig. 1c). This corresponded to a high 14.7-fold increase over the control uvr2 plants with 4.4 mutations per genome and generation (Fisher's exact test, P < 2.2E − 16). The 7.3 new mutations in UV-B-treated uvr3 plants represented a lower, but still significant 2.1-fold increase over the control treatment (Fisher's exact test,
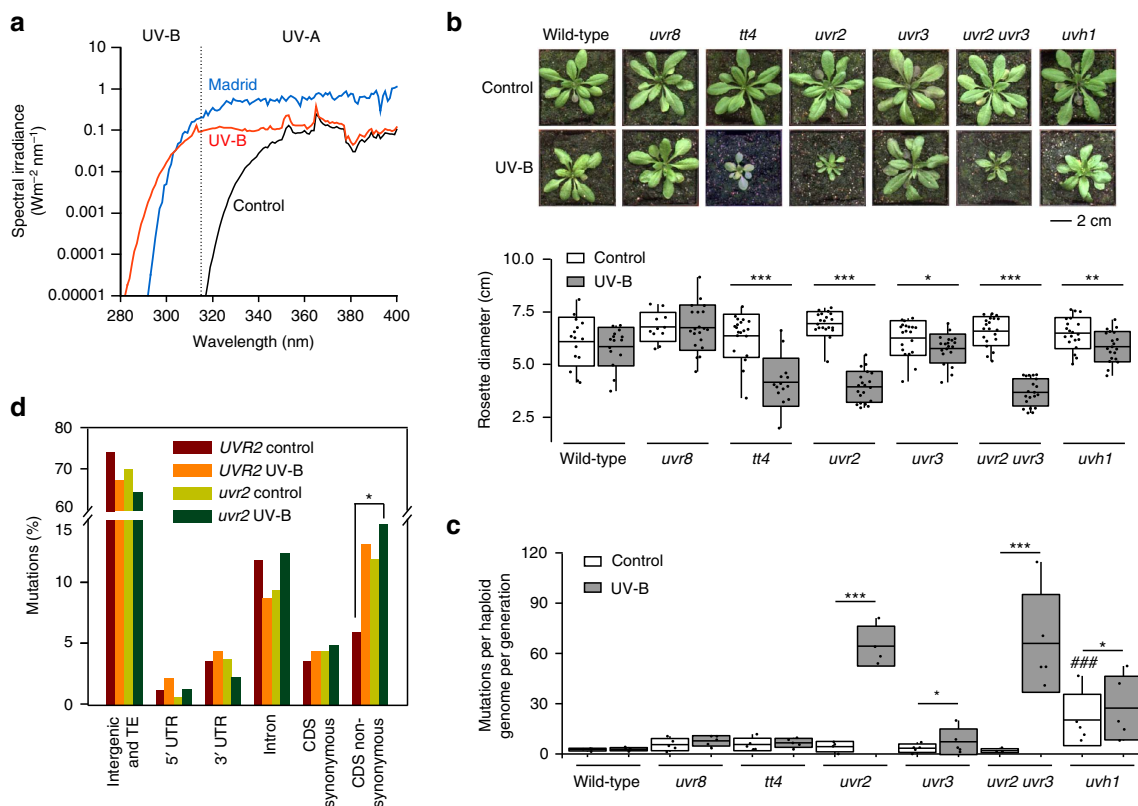
**Figure 1 | Frequencies and spectra of UV-B treatment-induced mutations. (a)** Spectral irradiance in sun simulator of the UV-B-free control (black; UV-B$_{BE}$ normalized at 300 nm (ref. 21) = 0 mW m$^{-2}$), and the simulated UV-B level of Madrid (red; UV-B$_{BE}$ normalized at 300 nm (ref. 21) = 300 mW m$^{-2}$) in UV-B and UV-A range (divided by dotted vertical line). The modelled Madrid UV-B$_{BE}$ (blue; UV-B$_{BE}$ normalized at 300 nm (ref. 21) = 265 mW m$^{-2}$) was generated using the Quick Tropospheric UV Radiation Calculator. **(b)** Representative phenotypes of individual genotypes grown under control and 300 mW m$^{-2}$ UV-B$_{BE}$. Rosette diameter measurements were performed on 11–20 plants per genotype and treatment. Significant differences in Student's $t$-test (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). **(c)** Normalized number of control and mutations induced by UV-B treatment per haploid genome and generation. Boxes show genotype average (middle line), s.d. (lower and upper margins), and values outside of the s.d. range (vertical bars). Dots represent individual genomes. ### and * indicate statistically significant (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$) differences in Fisher's exact test between: ### mutants versus wild-type and *control versus UV-B treatments (300 mW m$^{-2}$ UV-B$_{BE}$) of the same genotype, respectively. **(d)** Frequency of non-synonymous amino-acid changes in different genomic regions. UVR2 includes Col-0, uvr8, tt4 and uvr3 genotypes treated with 0 mW m$^{-2}$ UV-B$_{BE}$ (control) or with 100, 150 and 300 mW m$^{-2}$ UV-B$_{BE}$ (UV-B). uvr2 includes uvr2 and uvr2 uvr3 genotypes treated as control and UV-B. Numeric values are provided in Supplementary Table 1. *statistically significant (*$P = 0.0254$) difference in Fisher's exact test. All other comparisons within groups were not significant.

$P = 0.01965$). UV-B-exposed uvr2 uvr3 double-mutant plants had 66.0 new mutations (Fisher's exact test, $P \leq 2.2\text{E} - 16$; Fig. 1c). The progeny of uvr2 uvr3 plants exposed to 0, 100, 150 and 300 mW m$^{-2}$ UV-B$_{BE}$ revealed on average 2.0, 39.1, 65.3 and 66.0 mutations per haploid genome and generation, respectively (Supplementary Fig. 3b). This corresponded to 19.5-, 32.6- and 33-fold increase and indicated a UV-B dose-dependent accumulation of mutations at the lower and saturation at the higher UV-B doses, respectively (Fisher's exact test; all $P < 2.2\text{E} - 16$ in UV-B versus control; UV-B$_{BE}$ of 100 versus 150 and 300 mW m$^{-2}$: $P = 2.0\text{E} - 08$ and $1.2\text{E} - 08$; UV-B$_{BE}$ of 150 versus 300 mW m$^{-2}$: $P = 0.8978$).

The UV-B treatment also affected the frequency of non-synonymous amino-acid mutations. They were approximately threefold more frequent in UV-B-treated (300 mW m$^{-2}$ UV-B$_{BE}$) uvr2 versus control wild-type plants (14.7% versus 5.9% of all mutations, respectively; Fisher's exact test $P = 0.0254$; Fig. 1d). In absolute terms, this corresponded to 10.2 new non-synonymous amino-acid mutations per one uvr2 plant, compared with an average of 0.2, 0.4 and 0.5 such mutations in control wild-type, control uvr2 and UV-B-treated wild-type plants, respectively (Supplementary Table 1). We also found phenotypically distinct plants in the third UV-B-irradiated generation of the

double mutant (see example of semidominant mutant in Supplementary Fig. 3c), suggesting an increased functional impact of the mutations induced by the UV-B treatment on gene integrity in UVR2-defective plants.

**Spontaneous and induced mutation spectra in *A. thaliana*.** To characterize the treatment-specific mutation spectra, we compared mutations from all control plants with those of all UV-B-treated plants with exception of uvh1 samples, which were excluded owing to a 35% rate of A:T→T:A transversions, compared with <10% in the other genotypes (Supplementary Fig. 4a).

Consistent with previous observation of Ossowski et al.[22], about half (52%) of all substitutions under UV-B-free conditions were G:C→A:T nucleotide transitions (Fig. 2a). The G:C→A:T frequency increased to 88% after UV-B treatment (Fisher's exact test $P < 2.2\text{E} - 16$), which led to significantly reduced proportion of all other substitution types (Fig. 2a; Fisher's exact test $P$ values for control versus UV-B; A:T→G:C, 2.0E − 02; A:T→T:A, 9.6E − 05; G:C→T:A, 2.1E − 05; A:T→C:G, 3.9E − 12; G:C→C:G, 1.3E − 03). Therefore, simulated natural UV-B caused almost exclusively G:C→A:T nucleotide transitions.

To test whether this holds true in major genome fractions, we quantified mutation spectra in genes and transposons separately (Supplementary Fig. 4b). Under control conditions, G:C→A:T nucleotide transitions remained the major type of change in transposons (66%); however, this trend was absent in genes (23%)

where all six possible substitution types showed relatively similar frequencies (10–23%). We also observed more G:C→A:T nucleotide transitions in transposons (65%) than in genes (42%) within the data of Ossowski et al.[22] (Supplementary Fig. 4c). Surprisingly, after UV-B treatment, the G:C→A:T transition rate changed and was even larger in genes than in transposons (93% versus 87%; Fisher's exact test, P value = 0.0038; Supplementary Fig. 4b). Hence, transposons were prone to G:C→A:T transitions under both control and UV-B conditions, while genes only during UV-B treatment.

To find whether spontaneous mutation and those induced by UV-B treatment occurred in a particular sequence context, we performed a motif analysis around mutated sites. This revealed an absence of any specific mutation-prone context in the vicinity of spontaneously mutated G:C→A:T sites in control samples (Fig. 2b). However, within UV-B-treated plants C→T and G→A mutations occurred preferentially within the TC(C/T) and (G/A)GA contexts, respectively. Such an asymmetric and reverse complementing pattern strongly suggests that: (i) G→A mutations are C→T mutations on the reverse strand; (ii) mutations induced by the UV-B treatment occur predominantly at the 3′ base of the pyrimidine dimer; and (iii) that TC(C/T) represents the UV-B-mutation-prone sequence in A. thaliana.

**DNA methylation overlaps with the mutated sites.** On the basis of the preferential UV-B mutagenesis of DNA-methylated cytosines in the CpG context in mammals[23,24], we tested for correlation between DNA methylation patterns and mutations induced by the UV-B treatment in A. thaliana. Because DNA methylation is a very stable epigenetic modification, we used existing genome-wide DNA methylation data sets[25,26]. According to the functional types of DNA methylation in plants[25], we classified cytosines in the CG, CHG and CHH sequence contexts (where H is A, T or C) as being either methylated or non-methylated and scored for the methylation status at mutated
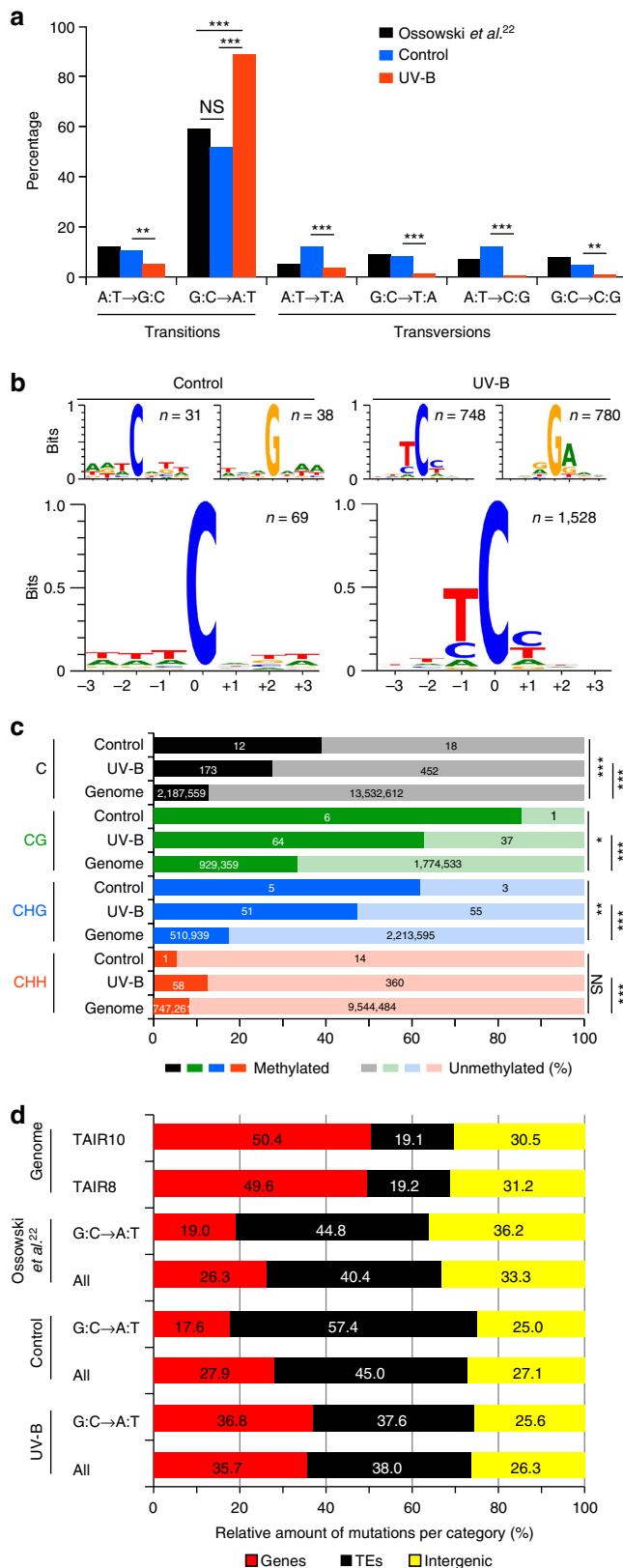


**Figure 2 | Genomic features of mutated positions.** (**a**) Proportions of single-nucleotide changes in Ossowski et al.[22] control samples (includes all genotypes treated with 0 mW m⁻² UV-B_BE; uvh1 was excluded) and UV-B-treated samples (includes all genotypes treated with 100, 150 and 300 mW m⁻² UV-B_BE; uvh1 was excluded). Statistical significance in Fisher's exact test: *P < 0.05, **P < 0.01, ***P < 0.001, n.s. = not significant. (**b**) DNA sequence motifs associated with control and mutations induced by UV-B treatment. Top images show cytosine and guanine mutation contexts on the forward strand. Bottom images show integrated information from both strands. Stacks' height indicates the sequence conservation measured in bits[44]. Symbol of mutated base at the position 0 was size reduced from 2 to 1 bit to reduce graph height. Height of other bases was not changed. Genomes are grouped into control and UV-B samples as described in **a**. (**c**) Percentage (x axis) of overlap of mutated positions with DNA methylation, and genome-wide DNA methylation frequencies for cytosines in C, CG, CHG and CHH contexts (where H is A, T or C). Values in columns show absolute number of mutated (Control and UV-B) or genomic positions (Genome) with available DNA methylation information. Statistical significance in Chi-square test with Yates correction: *P < 0.05, **P < 0.01, ***P < 0.001, n.s. = not significant. None of the control versus UV-B comparisons was significantly different (P > 0.05). Control samples were grouped as described in **a**. UV-B contained also 300 mW m⁻² UV-B_BE samples. (**d**) Percentage of mutations in major genome fractions. A. thaliana genome composition according to TAIR8 and TAIR10 annotations. Proportions of spontaneous (Ossowski et al.)[22] control sun simulator and UV-B-treatment-induced (300 mW m⁻² UV-B_BE) mutations in genes, transposable elements (TE) and intergenic regions. Groups were analysed as 'all' mutations and G:C→A:T mutations only. Individual genotypes were grouped into control and UV-B samples as described in **a**.

positions. This revealed that both spontaneous and induced mutations overlapped with methyl-cytosines (with the exception of the CHH control group, which contained only 15 testable positions) significantly more often than expected at random based on the genome-wide DNA methylation frequencies (Chi-square test with Yates correction, $P$ values for control versus genome and UV-B versus genome: CNN: $1.12E-04$ and $<2.2E-16$; CG: $1.38E-02$ and $<2.2E-16$; CHG: $6.59E-03$ and $<2.2E-16$; CHH: $6.83E-01$ and $3.10E-07$; Fig. 2c). Hence, this suggests that methyl-cytosine is prone to mutate under UV-B conditions compared with non-methylated cytosine.

Because DNA methylation is concentrated into transposon-rich chromosomal regions in *A. thaliana*[25,26], we tested whether the mutations show particular genomic distribution. Both control and UV-B treatments led to hypo-accumulation of mutations in genes, relatively random accumulation in intergenic regions and hyper-accumulation in transposons (Fig. 2d). We confirmed this trend using independent data set of Ossowski *et al.*[22] However, UV-B treatment induced ~10% more mutations in genic regions compared with control plants. Therefore, the UV-B treatment adds to the mutagenic effect of DNA methylation, but also affects non-methylated cytosines in genic regions.

**Accumulation of induced mutations during development**. Early embryonic separation of gametic and somatic cell lineages largely prevents transgenerational inheritance of somatic mutations in mammals[27]. In contrast, the late separation of germline cells in plants[28] allows the inheritance of mutations induced during vegetative growth in cells of the apical meristem into the progeny. Alternatively, mutations can occur later after separation of male and female cell lineages and/or gamete formation. To determine whether mutation induced by UV-B treatment accumulated during particular developmental stages, we analysed the ratio of heterozygous and homozygous mutations in the progeny of the first generation of plants in control and UV-B treatments. If all mutations occurred before the differentiation of the male and female organs, we expected a 2:1 ratio of heterozygous versus homozygous mutations in an inbreeding constitutively monoecious species such as *A. thaliana*. We found ratios of 1.4:1 (wild-type control), 2.5:1 (wild-type UV-B-treated) and 1:1 (*uvr2* control), but there were significantly 8.1-fold more heterozygous than homozygous mutations (44.22 versus 5.44 per haploid genome, respectively) in the progeny of UV-B-treated *uvr2* plants (Fisher's exact test $P$ values when compared with the other groups: $2.95E-08$, $5.83E-05$ and $7.97E-05$, respectively; Fig. 3a). This strongly suggested that the combination of UV-B treatment with *uvr2* genotype leads to mutations mostly after the split of female and male cell lineages. To validate this, we expressed luciferase-tagged UVR2 under control of its native promoter (*UVR2promoter::UVR2:LUCIFERASE*). The reporter line showed strong UV-B-independent developmentally controlled UVR2 accumulation in meristems (root apical meristem, young leaves, flowers, flower buds, axillary buds, closed anthers and young pistils), scars after petals and sepals and weaker expression in expanded leaves (Fig. 3b–e; the control non-transgenic plants are shown in Supplementary Fig. 5). No expression was observed in green or dry seeds (Fig. 3e). The strong UVR2 expression in floral tissues supported the results of our genetic analysis.

Occurrence of a high number of mutations in male and female cell lineages allowed us to test whether there are sex-specific preferences in mutation accumulation in *A. thaliana*. We grew *uvr2 uvr3* plants under control UV-B-free conditions until bolting, and then exposed half of the plants to UV-B until flowering and subsequently reciprocally crossed UV-B-irradiated and control plants (Fig. 3f). The resulting F1 hybrids were grown under non-UV-B conditions, and genomes of eight plants per crossing
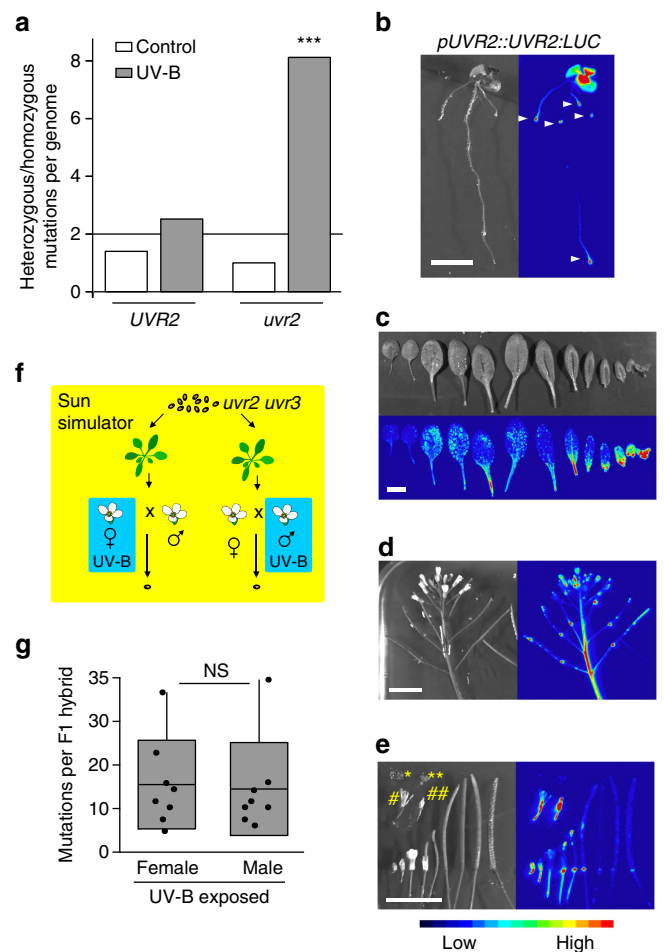


**Figure 3 | Developmental aspects of mutagenesis by UV-B treatment.**
(**a**) Ratio of heterozygous versus homozygous mutations in UVR2 (wild-type, *uvr8*, *tt4* and *uvr3*) and *uvr2* (*uvr2* and *uvr2 uvr3*) genotypes after one generation of control and UV-B treatment ($300\,\mathrm{mW\,m}^{-2}$ UV-B$_{BE}$). The 2:1 ratio (horizontal line) was expected if all inherited mutations occurred during somatic development. Mutations above this ratio were likely to originate after separation of male and female cell lineages. *** indicates statistically significant differences to all other samples in Fisher's exact test, $P<0.001$. (**b**–**e**) Expression of UVR2-LUCIFERASE translational fusion construct driven by endogenous promoter (*UVR2promoter::UVR2:LUCIFERASE*). Images on the top/left show plant tissues under white light and those on the bottom/right luciferase signal. All luciferase images were taken using identical exposure time of 1 min, and colour scale at the bottom indicates signal intensity. (**b**) Ten-day-old *in vitro* grown plant. Arrowheads indicate luciferase signals in root apical meristems. Scale bar, 5 mm. (**c**) Leaves dissected from 3-week-old *A. thaliana* plant organized from the oldest (left) to the youngest (right). Scale bar, 10 mm. (**d**) Inflorescence. Scale bar, 10 mm. (**e**) Flower, silique and seed developmental series. Bottom row, left to right: closed flower, flower with emerging pistil, fully opened flower, siliques at different stages and the last opened silique containing seeds with mature embryos. Hashes: pistils and anthers from (#) opened and (##) closed flowers. Petals and sepals were manually removed. Asterisks: (*) dry and (**) fresh seeds. Scale bar, 10 mm. (**f**) Genetic test for sex specificity of UV-B-induced mutations. *uvr2 uvr3* control and UV-B-irradiated plants ($300\,\mathrm{mW\,m}^{-2}$ UV-B$_{BE}$) were reciprocally crossed and the number of female- and male-specific mutations was analysed in progeny plants. (**g**) Boxes show genotype average (middle line), s.d. (left and right margins) and values outside of the s.d. range (horizontal bars) between eight analysed genomes (dots) per experimental point. NS, not significant (Student's *t*-test, $P=0.844$).

direction were sequenced and analysed. All recovered mutations were heterozygous, excluding self-pollination in any of the 16 analysed genomes (Supplementary Data 2). We found on average 12.4 mutations per UV-B-irradiated mother and 13.3 per UV-B-irradiated father, respectively (nonsignificant in Student's $t$-test, $P = 0.844$; Fig. 3g and Supplementary Table 2). This suggests that UVR2 is required for protection of both female and male genome stability, and UV-B treatment induces a similar number of mutations in both sexual lineages.

## Discussion

Land plants are exposed to solar UV-B during their entire life[3]. In order to minimize UV-B-induced damage, plants use multiple protection and repair pathways, including flavonoid sunscreen, direct reversal of pyrimidine dimers and NER[6,8,15,29,30]. We determined the frequency of transgenerationally inherited mutations induced by UV-B treatment in A. thaliana wild-type and mutant plants treated with simulated solar UV-B, resembling natural conditions from Helsinki (south Scandinavia) to Madrid (central Spain).

The simulated natural UV-B conditions had only a minimal effect on the rosette growth of wild-type Col-0, indicating that they were well in the photomorphogenic range. A wild-type-like phenotype of the UV-B photoreceptor mutant was unexpected as uvr8 was found to be UV-B-hypersensitive in previous studies[19,31,32]. The most likely reasons were acute UV-B stress doses applied to non-acclimated plants and/or use of mutants in more sensitive genetic background in the other studies. In contrast, tt4 and uvr2 plants were highly sensitive to the simulated natural UV-B, suggesting that flavonoid production and CPD repair, respectively[6,13], are the most important mechanisms sustaining plant growth under simulated natural UV-B.

Under control conditions, we observed on average $2.3 \times 10^{-8}$ mutations per site, which is approximately threefold more than the previously estimated mutation rates of $7.1$–$7.4 \times 10^{-9}$ for A. thaliana[22,33]. This could be because of presence of UV-A and/or higher photosynthetically active radiation (PAR; 400–700 nm; 340 µmol m$^{-2}$ s$^{-1}$) fluence rate applied in our control treatment compared with a typical A. thaliana growth chamber (100–150 µmol m$^{-2}$ s$^{-1}$). However, PAR applied in this study corresponds to a partially shaded natural site, while the full exposure to the sun is simulated using much higher PAR fluence rates (800 µmol m$^{-2}$ s$^{-1}$; refs 11,19). Simulated natural UV-B conditions caused only small (1.2–2.2-fold) increase in mutation rates of Col-0 wild-type plants. This is in agreement with a previous study, where simulated solar UV-B regimes provoked only one to four germinal somatic homologous recombination events per 250,000 seedlings[11].

The robust protection of A. thaliana transgenerational genome stability against UV-B strongly depends on direct reversal by UVR2 CPD photolyase (summarized as schematic model in Fig. 4). The uvr2 plants accumulated, on average, 64.3 new mutations per haploid genome and generation under the simulated central Spain UV-B regime. Some of these mutations apparently led to a loss of function for housekeeping genes within just three generations. In contrast, loss of UVR3 and UVH1 resulted in a significant, but much lower number of mutations. This may reflect low abundance of UV-B-induced (6–4)PPs (10–25%) relative to CPDs (75–90%) and partial redundancy of NER and UVR3 in repair of (6–4)PPs but not CPDs in A. thaliana[13,29].

DNA sequences prone to accumulate UV-B-induced mutations have been unknown in plants. We showed here that sensitivity to our UV-B treatment is determined by both genetic and epigenetic means. Mutations occurred preferentially in the TC dipyrimidine sequence context, and were enriched at methylated cytosines. This differed from spontaneous mutations, which were determined mainly epigenetically by DNA-methylated sites in transposons, but showed no association with particular short sequence motifs. The typical A. thaliana-hypermutable sequence TC(C/T) identified here differed from those in humans in at least two aspects. First, we did not observe any CC to TT dinucleotide mutations, which were found frequently in the human eyelid cells[34]. Second, in human skin cells the mutated cytosine was frequently followed by a guanine ((T/C)CG)[23]. A high proportion of (T/C)CG mutations in humans is most likely caused by the enhanced formation of pyrimidine dimers at methylated cytosines[23,24,35,36], which are found exclusively in the CG context in mammalian somatic cells[37]. Absence of such pattern in A. thaliana can be explained by presence of DNA methylation in any cytosine context in plants and low number of methylated cytosines in the A. thaliana genome[25,26]. Although mutations induced by our UV-B treatment were enriched in A. thaliana at the positions of methyl-cytosines (27%) relative to genome background (15%), they were not limited to them, and majority of the mutations (73%) appeared at non-methylated positions. This trend was weaker for spontaneous mutations (60% at non-methylated sites) and suggested that UV-B and spontaneous mutations may quantitatively differ in generating C→T transitions via indirect (involving uracil intermediate) or direct conversion, respectively[38].

Animal male and female germline cells separate from somatic cell lineages early during embryo development, and the latter do not divide any more during the post-embryonic phase[39]. In contrast, plant germline cells with undifferentiated sex divide several times during vegetative growth and separate into male- and female-specific cell lineages only during late flower development[40]. This potentially increases the risk of
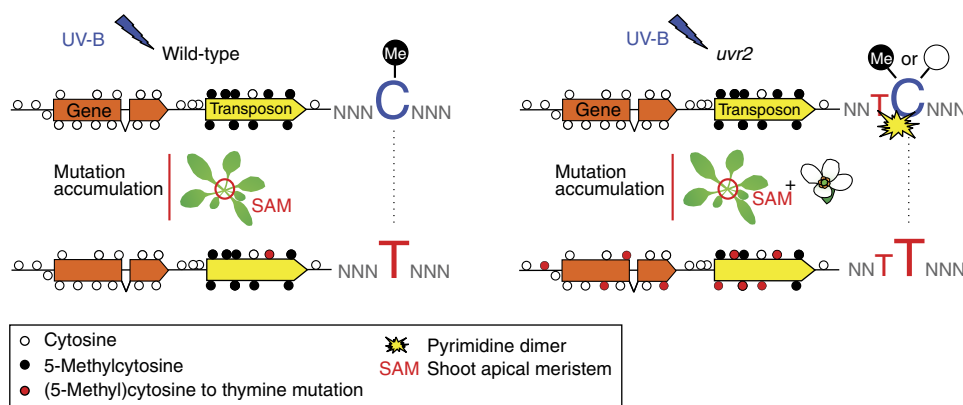


**Figure 4 | Model for accumulation of UV-B-induced germline mutations in A. thaliana.**

inheriting mutations via somaclonal sectors. In the first post-irradiated generation of control and UV-B-irradiated plants, we found ∼1:2 ratios of homozygous and heterozygous mutations, respectively. This showed that the spontaneous mutations occurred before the split of male and female cell lineages and the same was true also for mutations induced by UV-B treatment in *UVR2* plants. However, there were fourfold more heterozygous mutations in progenies of UV-B-irradiated *uvr2* plants. This provided strong genetic evidence that UVR2 prevents UV-B-induced mutations in germline cells mainly after separation of male and female cell lineages, and this UVR2 function seems complementary to its role in resolving CPDs in somatic cells[13]. In mammals, mutation rates can be much higher in male than in female gametes[39]. Here we showed that *uvr2* plants derived from UV-B-irradiated male and female reproductive tissues carry almost identical numbers of mutations, suggesting that male and female mutation rates may be more equal in plants. Mammalian mutation bias is caused by accumulation of mutations from DNA replication errors in sperms, which are products of many more cell generations than eggs[39]. It is unknown how many cell divisions (and DNA replications) are required for the development of *A. thaliana* anthers and carpels; however, the information is available from meiosis onwards. At the onset of meiosis there is a single round of DNA replication followed by two rounds of cell division. Subsequently, the released microspore undergoes two rounds of DNA replication and cell division resulting in one vegetative and two sperm cells. The megaspore replicates and divides three times and produces embryo sac with seven nuclei, including haploid egg cell[41]. Hence, there is comparable number of DNA replications in plant mega- versus microgametogenesis. This may explain similar number of mutations observed in our experiments; however, on the other hand it also shows that CPD direct reversal is important in both *A. thaliana* sexual lineages. This is unexpected because eggs are embedded much more in plant tissues than pollen and, therefore, should receive less UV-B damage. We speculate that this may be due to greatly reduced (haploid and unreplicated) genome constitution during gametogenesis, which may limit availability of homologous chromosomes and sister chromatids for homology-based DNA damage repair.

In addition to its activity in somatic cells, direct reversal of CPDs by UVR2 is the key mechanism protecting integrity of DNA from UV-B-induced mutations in *A. thaliana* male and female germline tissues. Direct reversal activity may be particularly important during plant haploid stage, when homology-based repair pathways may not be fully effective because of limited template availability. Therefore, UVR2 is necessary to avoid solar UV-B-induced genetic defects that could be transmitted to the future generations.

## Methods

**Simulation of solar radiation.** Simulation of solar radiation was performed in the sun simulators of the Research Unit Environmental Simulation at the Helmholtz Zentrum München, Neuherberg, Germany. Simulated spectra (280–850 nm; Fig. 1a and Supplementary Fig. 1a,b) were obtained by a combination of metal halide lamps (HQI/D, 400 W; Osram, München, Germany), quartz halogen lamps (Halostar, 300 and 500 W; Osram), blue fluorescent (TLD 18, 36 W; Philips, Amsterdam, the Netherlands) and UV-B fluorescent tubes (TL12, 40 W; Philips). The natural balance from ultraviolet to infrared radiation was achieved by filtering through borosilicate, lime and acrylic glass filters and a water layer and measured using a double monochromator system (Bentham, UK). The filtering in control condition excluded the entire UV-B, present in UV-B treatments. Owing to filter characteristics, ∼80% and more of UV-A were transmitted at control conditions for wavelength >360 nm compared with UV-B treatments, whereas at shorter wavelength of 330 nm only 10% were transmitted (Supplementary Fig. 1a,b). The standard growth conditions were set to resemble the main *A. thaliana*-growing season: day = 14 h, 21 °C, relative humidity 60%, PAR = 340 μmol m$^{-2}$ s$^{-1}$, which resembles natural PAR at shady sites; night = 10 h, 16 °C, relative humidity 80%, no PAR, UV-B radiation 1 h after onset of PAR for 10 h. Dusk and dawn was simulated by switching on/off different

groups of lamps. Four irradiation conditions were applied corresponding to: 0 (control), 100, 150 and 300 mW m$^{-2}$ UV-B$_{BE}$ normalized at 300 nm according to the generalized plant action spectrum[21] (Fig. 1a and Supplementary Fig. 1b). This realistically mimics UV-B$_{BE}$ doses during spring in northern mid-latitudes (40°N, 50°N, 60°N) at, for example, Madrid, Berlin and Helsinki, respectively. The simulated UV-B$_{BE}$ (ref. 21) dose of 300 mW m$^{-2}$ (ultraviolet index = 6; UV-B = 1.2 W m$^{-2}$), applied widely in this study, matched well the integrated values of the spectral irradiance in Madrid (UV-B$_{BE}$ (ref. 21) = 265 mW m$^{-2}$; ultraviolet index = 7; UV-B = 1.3 W m$^{-2}$; modelled for 30 March 2015, 12:00 GMT (total ozone column of 300 DU, surface albedo of 0.1), using the Tropospheric Ultraviolet and Visible model; http://cprm.acom.ucar.edu/Models/TUV/Interactive_TUV/; Fig. 1a).

**Plant material.** Following *A. thaliana* homozygous genotypes in Col-0 background were used: wild-type; *uvr8-6* null[19] (SALK_033468), *tt4* (SALK_020583C), *uvh1* (SALK_096156C), *uvr2* (WiscDsLox466C12), *uvr3* (WiscDsLox334H05) and *uvr2 uvr3*. Each genotype was amplified twice by a single seed descent to reduce any potential heterozygosity, and the resulting seed population was bulk-genotyped before mutation accumulation experiments (Supplementary Fig. 1c). Seeds were sown on a standard soil, and 15 plants per genotype were kept in the described UV-B conditions until seed harvest. Using a single seed descent amplification strategy, we produced three UV-B-irradiated generations (Supplementary Fig. 1c). Note that the sequenced and the irradiated plants were not identical, but siblings (that is, seeds from G1 UV-B-irradiated patent were split into several parts. One part was grown in sun simulator as UV-B-irradiated G2 and the second part was grown in non-UV-B chamber to obtain material for sequencing). This was done in order to avoid stressing UV-B-irradiated plants by additional wounding damage that could potentially influence mutation frequencies.

The *UVR2promoter::UVR2:LUCIFERASE* reporter line was constructed using the Gateway System (Invitrogen) and the Gateway binary vector pGWB435 was used to fuse firefly's *LUCIFERASE* gene to the C terminus of UVR2. The line was stably expressing the construct over multiple generations and T-DNA was excluded to disrupt a gene open reading frame by mapping T-DNA position using TAIL-PCR.

**Nucleic acid isolation and whole-genome sequencing.** From 15 irradiated plants per generation, genotype and treatment, we selected randomly five individuals and grew one progeny plant per individual in a chamber without UV-B radiation for 3 weeks. Subsequently, vegetative rosettes were harvested and DNA extracted with a Nucleon Phytopure Kit (GE Healthcare). Sequencing libraries were prepared using a TruSeq DNA Kit (Illumina). Fragment sizes and library concentrations were assessed on a Bioanalyzer (Agilent) and high-quality libraries were 100 bp paired-end-sequenced on a HiSeq2500 (Illumina) instrument to an average 35 × genome coverage (Supplementary Fig. 2 and Supplementary Data 1).

**Mutation detection and validation.** Reads were adaptor- and quality-trimmed using SHORE (v8; ref. 42). Filtered and trimmed reads where aligned to Col-0 reference sequence (TAIR10, 119 Mbp) using GenomeMapper[43] integrated in SHORE (v8) using a maximum of 5% of the read length as mismatches including a maximum of 5% gaps. Read pair information was used to help to remove redundant alignments. Only uniquely mapped reads (after read pair correction) were considered. In order to remove reads originating from the same molecule (because of PCR amplification), we also removed reads with identical 5′ alignments using SHORE. Next, we generated a genome matrix containing information on total coverage and the single base counts for A,C,G,T,- and N for each re-sequenced genome at each reference sequence position. Positions covered by <20 reads were marked as low coverage. All other positions were classified as follows: (i) homozygous wild-type, (ii) homozygous mutant, (iii) heterozygous or (iv) undefined based on the allele frequency of the non-reference alleles. Frequency thresholds were determined empirically (Supplementary Data 1 and Supplementary Figs 6 and 7). Low complexity and tandem repetitive genome regions (comprising 2.95 Mb of the reference sequence), identified by RepeatMasker and TandemRepeatFinder, were excluded during this step to avoid false-positive mutation calls.

Novel mutations should be specific to the genome under consideration (focal genome). Therefore, we compared the variant/allele call in the focal genome with the alleles in nine other genomes of the same genotype (using only the first generation). For focal genomes in generations two and three, we excluded the respective parental genome from this filtering step. A variant call was considered as novel mutation, if none of the other nine genomes showed the same variant and at least six of them showed evidence for a homozygous wild-type allele at this position (Table 1). In addition, we used the following criteria for background filtering: (i) more than one of the background genomes is labelled 'undefined'; (ii) one of the background genomes shows a different homozygous or heterozygous mutation at the same base position; (iii) more than three of the background genomes are insufficiently (<20 ×) covered; or (iv) less than six background genomes have homozygous wild-type allele calls at the respective position.

We kept track of each position that could be analysed in the focal sample even if the position was called homozygous wild-type (accessible sites), in order to assess the frequency of mutated versus non-mutated accessible sites. The accessible sites

## Table 1 | Mutation classification thresholds.

| Frequency | Classification |
|---|---|
| >0.9 | Homozygous mutation, accepted |
| 0.8–0.9 | Undefined, mutation not accepted |
| 0.3–0.8 | Heterozygous mutation, accepted |
| 0.1–0.3 | Putative sequencing error, not accepted |
| <0.1 | Reference allele, accepted |

included $\sim$75% of the $\sim$120 million sites of the nuclear genomes. Normalized number of mutations per genome was calculated as $n$, where: $n = ((\text{total genome}/\text{accessible genome}) \times \text{number of accepted mutations})/\text{number of treated generations}$. Assignment of mutations to different genome regions (genes, TEs and intergenic regions) was carried out using current *A. thaliana* genome annotations (TAIR10) for genes and TEs. If a TE overlapped with a gene model, we considered the overlapping part as TE, based on the notion that this is frequently DNA-methylated in all cytosine contexts. TE genes were also treated as TEs in our analysis.

**Estimation of false mutation rates with simulated data.** We introduced 900 *in silico* mutations into the Col-0 reference sequence (TAIR10); 308 were homozygous and 592 were heterozygous reflecting the spectrum of mutations reported in this study. We simulated 25 Mio 100 bp Illumina read pairs with an insert size of 370 bp and a sequencing error rate of 2% using wgsim (https://github.com/lh3/wgsim). The sequencing depth for the simulated genome was $41\times$, which is even slightly lower than the average coverage obtained for the real data ($60\times$). The analysis was performed as described before, and nine of the sequenced G1 Col-0 (five control and four Madrid-like UV-B) genomes were used for filtering as background genomes.

The allele frequency distribution for variable sites in the simulated genomes was similar to the distributions observed in real data (Supplementary Figs 6 and 7). However, as the simulated data showed many more variable sites, the simulated sequencing error rate (2%) appeared to be higher than in real data. We found a clear separation in allele frequencies of homozygous and heterozygous variants (Supplementary Figs 6 and 7b). However, the distribution revealed that many of the putative heterozygous variants with an allele frequency between 0.1 and 0.2 are masked by a huge amount of putatively erroneous sites with low mutant allele frequencies. In contrast, only a much smaller number of putative heterozygous sites was observed with an allele frequency between 0.2 and 0.8 in both data sets (Supplementary Figs 6 and 7a). Assuming that the frequencies of real heterozygous sites should be normally distributed with a mean of 0.5 implies that variants with a frequency <0.3 seemingly include a lot of false-positives. The minimal turning point at 0.3 in histogram indicates that using this as a cutoff ensures that we exclude the majority of false-positives while sacrificing only a very small number of true-positives. We found in total 91,500,586 (75% of the genome) accessible sites in the simulated data, which is similar to the real data. In all, 24% of the simulated mutations were in regions that were not accessible according to our definitions. Note that this does not affect the mutation rate estimations as mutation frequency is estimated across the number of accessible sites. Of the remaining 685 *in silico* mutations located at the accessible site, 684 were identified by our approach (Supplementary Fig. 7b). Only one heterozygous mutation could not be reported, as it had an allele frequency below 0.3. Together, this simulation suggests a false-negative rate of 0.15%. We did not encounter any false-positive in this simulation, suggesting that our strict cutoffs are very robust against false-positives even at high sequencing error rates. In order to support this finding, we tested a random set of 59 candidates from a total of 2,497 mutations identified in the real sequencing by Sanger sequencing. We were able to confirm 58 of them (Supplementary Data 2).

**DNA sequence motif analysis.** For each accepted mutation, we extracted positions three bases up- and downstream from the respective position. Mutations were grouped by the type of base change (for example, $C \rightarrow T$) and the extracted sequences were used as input for the software weblogo v3.4 (ref. 44), which generates bit scores for each base (A, C, G or T) at a specific position. If a base is found more often than expected according to the background probability of each base (here $C = G = 0.2$, $A = T = 0.3$), it gets a higher bit score.

**DNA methylation analysis.** DNA methylation data were retrieved from publicly available wild-type *A. thaliana* data sets GSM980986, GSM980987 and GSM938370 (ref. 26). Only nucleotide positions with $\geq 10$ sequencing reads were considered for analysis. A cytosine was considered as methylated if its methylation frequency reached $\geq 10\%$ in at least two biological replicates. Because these criteria are partially different from those applied in other studies[25,26], we obtained generally higher DNA methylation frequencies. Statistical significance of the results was tested as the number of methylated and unmethylated cytosines in sample A versus sample B using Chi-square test with Yates correction.

## References

1. Kami, C., Lorrain, S., Hornitschek, P. & Fankhauser, C. in *Current Topics in Developmental Biology* (ed. Marja, C. P. T.) 29–66 (Academic Press, 2010).
2. Heijde, M. & Ulm, R. UV-B photoreceptor-mediated signalling in plants. *Trends Plant. Sci.* **17**, 230–237 (2012).
3. Jenkins, G. I. Signal transduction in responses to UV-B radiation. *Annu. Rev. Plant Biol.* **60**, 407–431 (2009).
4. Ulm, R. *et al.* Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of *Arabidopsis. Proc. Natl Acad. Sci. USA* **101**, 1397–1402 (2004).
5. Rizzini, L. *et al.* Perception of UV-B by the *Arabidopsis* UVR8 protein. *Science* **332**, 103–106 (2011).
6. Li, J., Ou-Lee, T. M., Raba, R., Amundson, R. G. & Last, R. L. *Arabidopsis* flavonoid mutants are hypersensitive to UV-B irradiation. *Plant Cell* **5**, 171–179 (1993).
7. Piofczyk, T., Jeena, G. & Pecinka, A. *Arabidopsis thaliana* natural variation reveals connections between UV radiation stress and plant pathogen-like defense responses. *Plant Physiol. Biochem.* **93**, 34–43 (2015).
8. Hideg, É., Jansen, M. A. K. & Strid, Å. UV-B exposure, ROS, and stress: inseparable companions or loosely linked associates? *Trends Plant Sci.* **18**, 107–115 (2013).
9. Molinier, J., Ries, G., Bonhoeffer, S. & Hohn, B. Interchromatid and inter-homolog recombination in *Arabidopsis thaliana. Plant Cell* **16**, 342–352 (2004).
10. Pecinka, A. *et al.* Transgenerational stress memory is not a general response in *Arabidopsis. PLoS ONE* **4**, e5202 (2009).
11. Ries, G. *et al.* Elevated UV-B radiation reduces genome stability in plants. *Nature* **406**, 98–101 (2000).
12. Kimura, S. & Sakaguchi, K. DNA repair in plants. *Chem. Rev.* **106**, 753–766 (2006).
13. Jiang, C.-Z., Yee, J., Mitchell, D. L. & Britt, A. B. Photorepair mutants of *Arabidopsis. Proc. Natl Acad. Sci. USA* **94**, 7441–7445 (1997).
14. Radziejwoski, A. *et al.* Atypical E2F activity coordinates PHR1 photolyase gene transcription with endoreduplication onset. *EMBO J.* **30**, 355–363 (2011).
15. Liu, Z., Hossain, G. S., Islas-Osuna, M. A., Mitchell, D. L. & Mount, D. W. Repair of UV damage in plants by nucleotide excision repair: *Arabidopsis* UVH1 DNA repair gene is a homolog of *Saccharomyces cerevisiae* Rad1. *Plant J.* **21**, 519–528 (2000).
16. Schärer, O. D. Nucleotide excision repair in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **5**, a012609 (2013).
17. Mazza, C. A. *et al.* Functional significance and induction by solar radiation of ultraviolet-absorbing sunscreens in field-grown soybean crops. *Plant Physiol.* **122**, 117–126 (2000).
18. Castells, E. *et al.* det1-1-induced UV-C hyposensitivity through *UVR3* and *PHR1* photolyase gene over-expression. *Plant J.* **63**, 392–404 (2010).
19. Favory, J. J. *et al.* Interaction of COP1 and UVR8 regulates UV-B-induced photomorphogenesis and stress acclimation in *Arabidopsis. EMBO J.* **28**, 591–601 (2009).
20. Thiel, S. *et al.* A phytotron for plant stress research: How far can artificial lighting compare to natural sunlight? *J. Plant Physiol.* **148**, 456–463 (1996).
21. Caldwell, M. *Solar UV Irradiation and the Growth and Development of Higher Plants* Vol. 7 (Academic Press, 1971).
22. Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana. Science* **327**, 92–94 (2010).
23. Ikehata, H. *et al.* Mutation spectrum in UVB-exposed skin epidermis of Xpa-knockout mice: frequent recovery of triplet mutations. *Environ. Mol. Mutagen* **48**, 1–13 (2007).
24. Rochette, P. J. *et al.* Influence of cytosine methylation on ultraviolet-induced cyclobutane pyrimidine dimer formation in genomic DNA. *Mutat. Res.* **665** 7–13 (2009).
25. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis. Cell* **133**, 523–536 (2008).
26. Stroud, H. *et al.* Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).
27. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).
28. Walbot, V. & Evans, M. M. S. Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.* **4**, 369–379 (2003).
29. Britt, A., Chen, J., Wykoff, D. & Mitchell, D. A UV-sensitive mutant of *Arabidopsis* defective in the repair of pyrimidine-pyrimidinone(6-4) dimers. *Science* **261**, 1571–1574 (1993).
30. Kimura, S. *et al.* DNA repair in higher plants; photoreactivation is the major DNA repair pathway in non-proliferating cells while excision repair (nucleotide excision repair and base excision repair) is active in proliferating cells. *Nucleic Acids Res.* **32**, 2760–2767 (2004).

31. Brown, B. A. *et al.* A UV-B-specific signaling component orchestrates plant UV protection. *Proc. Natl Acad. Sci. USA* **102,** 18225–18230 (2005).

32. Kliebenstein, D. J., Lim, J. E., Landry, L. G. & Last, R. L. *Arabidopsis* UVR8 regulates UV-B signal transduction and tolerance and contains sequence similarity to human regulator of chromatin condensation 1. *Plant Physiol.* **130,** 234–243 (2002).

33. Yang, S. *et al.* Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523,** 463–467 (2015).

34. Martincorena, I. *et al.* High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348,** 880–886 (2015).

35. You, Y.-H., Szabó, P. E. & Pfeifer, G. P. Cyclobutane pyrimidine dimers form preferentially at the major p53 mutational hotspot in UVB-induced mouse skin tumors. *Carcinogenesis* **21,** 2113–2117 (2000).

36. Tommasi, S., Denissenko, M. F. & Pfeifer, G. P. Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases. *Cancer Res.* **57,** 4727–4730 (1997).

37. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic reprogramming in plant and animal development. *Science* **330,** 622–627 (2010).

38. Franchini, D.-M., Schmitz, K.-M. & Petersen-Mahrt, S. K. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu. Rev. Genet.* **46,** 419–441 (2012).

39. Wilson Sayres, M. A. & Makova, K. D. Genome analyses substantiate male mutation bias in many species. *Bioessays* **33,** 938–945 (2011).

40. Meyerowitz, E. M. Plants compared to animals: the broadest comparative study of development. *Science* **295,** 1482–1485 (2002).

41. Yadegari, R. & Drews, G. N. Female gametophyte development. *Plant Cell* **16,** S133–S141 (2004).

42. Ossowski, S. *et al.* Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18,** 2024–2033 (2008).

43. Schneeberger, K. *et al.* Simultaneous alignment of short reads against multiple genomes. *Genome Biol.* **10,** 1–12 (2009).

44. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).

## Acknowledgements

## Author contributions

A.P., T.P. and K.S. designed the experiments. A.A. and J.B.W. planned and performed sun simulator experiments and determined growth parameters. T.P., A.A. and J.B.W. grew the plants. T.P. prepared sequencing libraries. E.-M.W. analysed whole-genome sequencing data and identified mutations. T.P. and E.-M.W. analysed mutation spectra and associations with genomic features. T.P. cloned the *UVR2* reporter construct. A.P. wrote the manuscript with contribution from all authors. All authors read and approved the submitted manuscript.

## Additional information

**Publikace 6**

Genome Biology

## RESEARCH

**Open Access**

CrossMark

# Recurrent evolution of heat-responsiveness in *Brassicaceae COPIA* elements

Björn Pietzenuk[1,2], Catarine Markus[1,3], Hervé Gaubert[4,5], Navratan Bagwan[1,6], Aldo Merotto[3], Etienne Bucher[7] and Ales Pecinka[1*]

## Abstract

**Background:** The mobilization of transposable elements (TEs) is suppressed by host genome defense mechanisms. Recent studies showed that the cis-regulatory region of Arabidopsis thaliana *COPIA78/ONSEN* retrotransposons contains heat-responsive elements (HREs), which cause their activation during heat stress. However, it remains unknown whether this is a common and potentially conserved trait and how it has evolved.

**Results:** We show that *ONSEN*, *COPIA37*, *TERESTRA*, and *ROMANIAT5* are the major families of heat-responsive TEs in *A. lyrata* and *A. thaliana*. Heat-responsiveness of *COPIA* families is correlated with the presence of putative high affinity heat shock factor binding HREs within their long terminal repeats in seven *Brassicaceae* species. The strong HRE of *ONSEN* is conserved over millions of years and has evolved by duplication of a proto-HRE sequence, which was already present early in the evolution of the *Brassicaceae*. However, HREs of most families are species-specific, and in *Boechera stricta*, the *ONSEN* HRE accumulated mutations and lost heat-responsiveness.

**Conclusions:** Gain of HREs does not always provide an ultimate selective advantage for TEs, but may increase the probability of their long-term survival during the co-evolution of hosts and genomic parasites.

**Keywords:** *Brassicaceae*, *COPIA*, Evolution, Heat stress, *ONSEN*

## Background

Transposable elements (TEs) are ubiquitous components of eukaryotic genomes. Their functions and roles range from DNA parasites, through regulators of gene transcription to facilitators of genome evolution (reviewed in [1, 2]). Together with other types of repeats, TEs comprise 10–80 % of plant genome content and specific families of long terminal repeat (LTR) retrotransposons can reach thousands of copies per genome [1, 3]. Plants evolved several layers of sophisticated epigenetic silencing mechanisms in order to suppress TE activity. Their transcripts are degraded by the post-transcriptional gene silencing (PTGS) pathway, which greatly reduces possible transposition events [4]. In parallel, transcriptional gene silencing (TGS) stably silences TEs by deposition of DNA methylation via RNA directed DNA methylation (RdDM) mechanism (reviewed in [5, 6]). The repressed state is further stabilized by accumulation of specific histone modifications and faithfully transmitted in a

DNA replication-dependent manner to the next generations. External or internal factors [7, 8] can lead to transient loss of silencing, but the epigenetic control will be re-established through tissue-specific RdDM activity [9]. In addition to the nimble epigenetic silencing system, entire TEs can be physically removed from the host genome by deletion-biased homologous recombination processes [10].

In spite of the multi-layer amplification barriers, many TE families show signs of recent transpositions [11–13], suggesting that TEs occasionally escape epigenetic surveillance. There is increasing evidence that stress treatments affect chromatin structure and may lead to transposon activation (reviewed in [1, 14]). A possible mechanism was proposed based on the analysis of stress-induced TEs. LTRs of medicago cold-inducible repetitive element (*MCIRE*) retrotransposon contain a putative cold-responsive element (CRE) in alfalfa (*Medicago sativa*) [15]. The CRE is specified by a conserved 5-bp core sequence (CCGAC) typical for C-repeat (CRT)/dehydration-responsive elements (DRE) that are recognized by cold-specific transcription factors (TFs) [16]. LTRs of heat-responsive *COPIA78/ONSEN* (used as synonyms in this study) retrotransposon in *Arabidopsis*

\* Correspondence: pecinka@mpipz.mpg.de
[1]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne 50829, Germany
Full list of author information is available at the end of the article

*thaliana* [7, 8, 17], contain a cluster of four nGAAn motifs forming a heat-responsive element (HRE) [18]. During heat stress (HS), the *ONSEN* HRE is bound by heat shock factor A 2 (HSFA2), which triggers its transcriptional activity. This regulation is very specific and greatly independent of TGS control as the loss of decreased DNA methylation 1 (*DDM1*) in mutant plants did not trigger *ONSEN* transcriptional activation [7], in contrast to other typical LTR retrotransposons [19].

Presence of HRE and CRT/DRE motifs in *ONSEN* and *MCIRE*, respectively, suggested that the TEs' response to stresses may be mediated by specific TF binding motifs. HREs were previously classified into four types based on their structure and, most likely, also activity [20]. The strongest 4P HRE contains at least four adjacent nGAAn motifs and is bound by two HSFA2 trimers. The 3P HRE is bound by a single HSFA2 trimer and represents a moderately responsive HRE. In contrast, gap and step HREs with irregularly and more distantly spaced nGAAn motifs have on average lower HRE activity. Therefore, the HRE composition needs to be considered in order to define the strength of transcriptional response.

Here we identified multiple heat-responsive *COPIA* families in *Arabidopsis lyrata* and *A. thaliana*, two closely related species, using RNA sequencing (RNA-seq). Subsequently, we extended our analysis to five other *Brassicaceae* species and reconstructed putative HREs, their evolutionary history, and validated our predictions by transcriptional analysis after HS treatment.

## Results

### Identification of heat-responsive TE families in *A. thaliana* and *A. lyrata*

First, we determined HS conditions that would be effective and comparable for *A. lyrata* MN47 and *A. thaliana* Col-0 plants. As the *A. lyrata* genome contains sequences with high homology to the *A. thaliana* *ONSEN* retrotransposon, we quantified *ONSEN* transcripts in both species by reverse transcription quantitative polymerase chain reaction (RT-qPCR) during a HS (37 °C) time series using soil-grown plants. Transcripts accumulated faster in *A. thaliana*, but to comparable amounts in both species after 12 h of HS (Fig. 1a). We selected 6 h at 37 °C, leading to a significant and reproducible *ONSEN* transcript accumulation in both species (T-test, $P$ <0.05), as the standard HS treatment. Subsequently, samples of control, heat-stressed (6 h HS), and recovered (6 h HS + 48 h 21 °C) plants were RNA-sequenced (Fig. 1b).

To assess the extent of plant responses to HS, we monitored transcript levels from 32,793 *A. lyrata* and 32,678 *A. thaliana* protein-coding genes. This revealed significant upregulation (adjusted $P$ <0.05; DESeq) of 21.8 % *A. lyrata* genes ($n$ = 7156) and 18.9 % *A. thaliana* genes ($n$ = 6165) after 6 h HS (Fig. 1c; Additional files 1 and 2). After

recovery, we found only 2.9 % ($n$ = 980) of genes still upregulated in *A. lyrata* and 0.6 % ($n$ = 192) in *A. thaliana*. *A. lyrata* showed 21.3 % ($n$ = 6992) downregulated genes after HS and 1.5 % (n = 491) after recovery (Fig. 1d). There were 17.3 % ($n$ = 5650) significantly downregulated genes after HS and only 0.3 % ($n$ = 89) after recovery in *A. thaliana*. Hence, HS treatment induced a similar degree of transient transcriptional changes in both species.

Because there is no publicly available *A. lyrata* TE annotation, we prepared custom-made catalogues of 53,089 *A. lyrata* and 17,009 *A. thaliana* repetitive elements (Additional files 3 and 4, respectively). Although the two species differed threefold in their TEs numbers, their spectra of TE families were similar (Additional file 5: Figure S1). The multi-copy nature of many TEs hinders RNA-seq analysis using standard protocols. Therefore, we developed the *COMparative EXpression of TEs* (COMEX) method, which allows quantification of transcripts derived from individual TE copies and effective removal of the RNA-seq reads mapping across TE families (see "Methods;" Additional files 6 and 7). We found 197 and 132 significantly (adjusted $P$ <0.05; DESeq) upregulated TEs, representing 90 and 60 families (26 in common), after 6 h HS in *A. lyrata* and *A. thaliana*, respectively (Fig. 1e, f; Additional files 8 and 9). Comparing the major upregulated TE groups versus those in the whole genome revealed general under-representation of *DNA transposons* and *HELITRONs* and *A. lyrata*-specific under-representation of SINEs. In contrast, we found an over-representation of heat-responsive *SADHU* and *LINE* retrotransposons in *A. lyrata*, *GYPSY* elements in *A. thaliana*, and *COPIA* TEs in both species (Fig. 1g). Heat-responsive *AlCOPIAs* (n = 60; 100 %) comprised six families with at least two heat-inducible elements (Fig. 1h): *AlCOPIA31* ($n$ = 3; 3 %), *AlCOPIA79* ($n$ = 2; 3 %), *AlCOPIA37* ($n$ = 5; 11 %), *AlCOPIA20* ($n$ = 9; 14 %), *AlONSEN* ($n$ = 19; 37 %), and a so far unknown family which we named **TE**MPERATURE **RES**PONSIVE **TRA**NSPOSON (*TERESTRA*, $n$ = 6; 10 %), as well as a bulk of single copies from different families ($n$ = 16; 22 %). *A. thaliana* heat-responsive *COPIAs* ($n$ = 34) were represented by six families with more than one heat-responsive TE. However, only *AtONSEN* ($n$ = 8; 29 %) and *AtCOPIA37* ($n$ = 4; 12 %) were common between both species (Fig. 1i). A prominent *A. thaliana*-specific family was *ROMANIAT5*, comprising 12 % ($n$ = 4) of all heat-responsive *AtCOPIAs*. After recovery, all TEs were re-silenced in *A. thaliana* and only five (*AlCOPIA37*, *AlRE1*, *SADHU6-1*, *AlATN9_1*, and *AlLINE1_3A*) showed increased transcript amounts in *A. lyrata* (Fig. 1e; Additional file 8). Surprisingly, *ONSEN* was fully silenced after two days of recovery, most likely owing to a shorter HS applied here compared to the previous study [7]. The families representing at least 10 % of heat-responsive *COPIA* elements in each species were considered for further analysis (Fig. 1h).
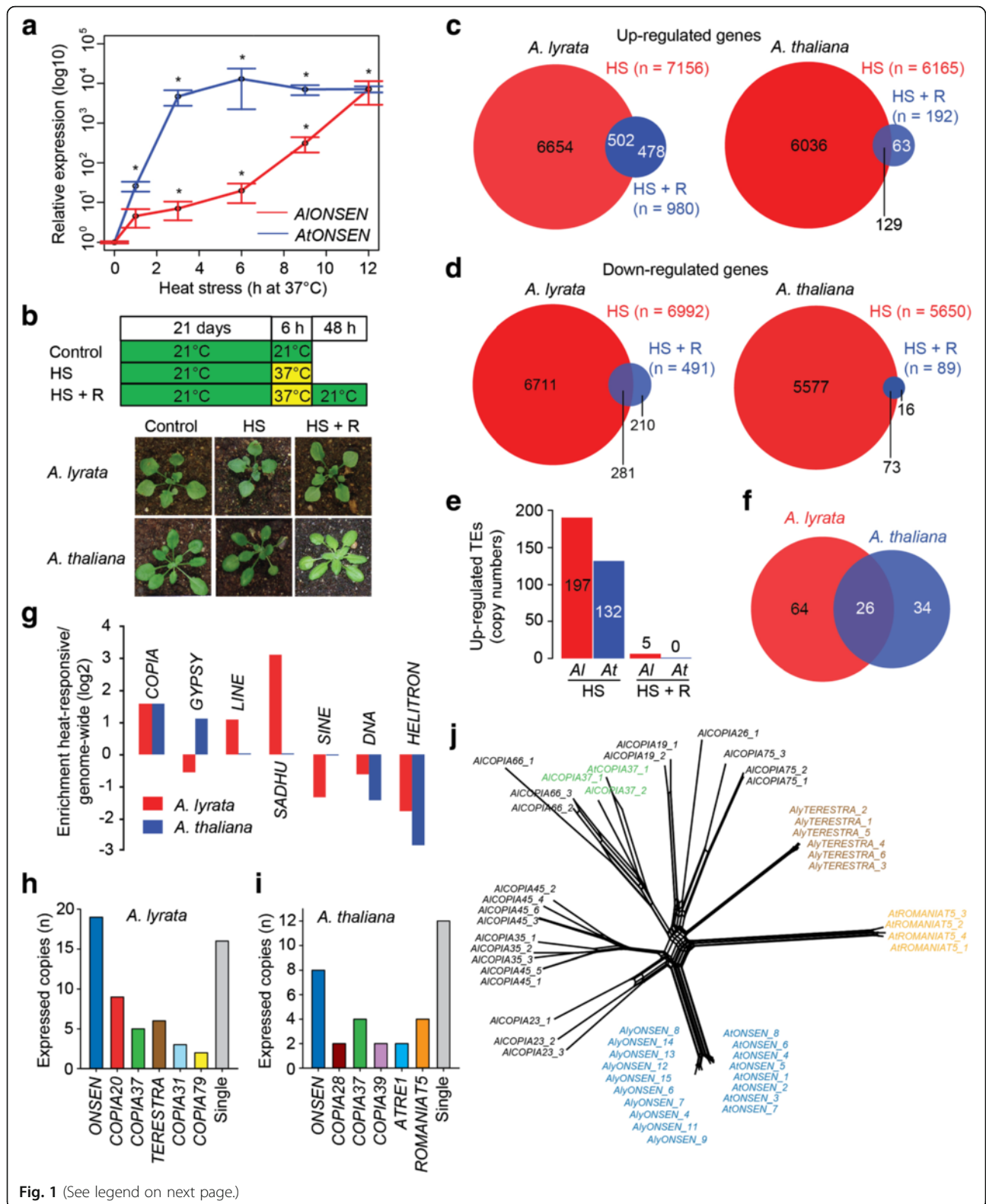
**Fig. 1** (See legend on next page.)

(See figure on previous page.)
**Fig. 1** Transcriptome analysis of heat-stressed *A. lyrata* and *A. thaliana* plants. **a** Effects of HS on *ONSEN* heat-responsiveness in *A. thaliana* and *A. lyrata*. Both species were stressed at 37 °C for the indicated number of hours (h) and subsequently analyzed for the amount of *ONSEN* transcript (log10) by RT-qPCR relative to *GAPD-H* transcript amounts. * significant (t-test, *P* <0.05) transcript enrichment relative to 0 h control. *Error bars* indicate standard deviation of three biological replicates. **b** Design of plant HS treatment for RNA-seq and representative phenotypes of control, 6 h heat-stressed at 37 °C and recovered plants. **c**, **d** Number of significantly (**c**) upregulated or (**d**) downregulated protein-coding genes after 6 h at 37 °C and 48 h recovery at non-stress conditions in both species. **e**, **f** Number of significantly upregulated (**e**) TEs and (**f**) TE families after 6 h HS and 48 h recovery. **g** Identification of TE groups enriched for heat-responsive copies. Retrotransposons were divided into *SINE, SADHU, LINE, COPIA,* and *GYPSY* family members. The relative enrichment of heat-activated TEs was calculated as ratio between % of all heat-activated to % of all TEs genome-wide and expressed on a log2 scale. The major heat-responsive *COPIA* families in (**h**) *A. lyrata* and (**i**) *A. thaliana*. The families containing a single HRE are displayed as "single copies." **j** RT amino acid sequences (Additional file 12)-based phylogenetic network of selected heat-responsive (*colored*) and non-responsive (*black*) *A. lyrata* and *A. thaliana COPIA* families. The data are also provided as un-rooted three in Additional file 5: Figure S2

Next, we tested whether heat-responsive *COPIA* families represent a particular *COPIA* clade. We reconstructed phylogeny of HS-responsive *COPIA37, ONSEN, TERES-TRA, ROMANIAT5,* and seven HS-non-responsive *COPIA* families (*19, 23, 26, 35, 45, 66, 75*) based on their RT sequences (Fig. 1j; Additional file 5: Figure S2). The coding sequence was preferred over LTRs for the similarity analysis because this is strongly influenced by length of the input sequences, which may vary drastically in case of LTRs from different families. All heat-responsive families formed distinct and early separated branches, suggesting multiple independent origins of *COPIA* heat-responsiveness.

### The structure and evolution of *ONSEN* heat-responsiveness

There are 24 *COPIA78* elements in *A. thaliana* Col-0 (TAIR10) including eight full-length copies and 16 fragments (Table 1, Additional file 5: Table S1). However, only the eight full-length *ONSEN* copies were found to be heat-responsive (Additional file 9). We performed in silico reconstruction of the putative HREs using a proposed classification [20], which suggested two HREs in all heat-responsive *AtONSENs*: a low efficiency gap HRE and the highest efficiency 4P HRE (Fig. 2a; Additional file 5: Figure S3). While the gap HRE is present in all eight *A. thaliana* full-length *ONSENs*, the 4P was changed into a 3P

HRE with moderate efficiency in *AtONSEN4*, due to loss of the fourth motif (Additional file 5: Figure S3). In contrast, none of the 16 fragments or solo LTRs contains functional HREs nor shows heat-responsiveness according to RNA-seq (Additional file 9).

We found 55 *COPIA78* TEs in *A. lyrata*. Ten are full-length elements and 45 are fragments, either solo LTRs or incomplete according to the gaps in the genome assembly (Table 1; Additional file 5: Figure S3 and Table S2). In total, 15 copies contain at least one putative HRE with three or more adjacent (≤5 bp) nGAAn motifs. Remarkably, a high number of *AlONSENs* carry HREs identical to *A. thaliana* copies (Fig. 2a; Additional file 5: Figure S3). *AlONSEN 2* and *8* have *A. thaliana*-like gap HREs; the 4P type is present in *AlONSEN 10* and both co-occur in *AlONSEN 6, 7, 9, 11, 14, 15, 17*. In addition, we observed putative low efficiency gap/step HREs substituting the 4P HRE in *AlON-SEN 1, 4, 5, 12,* and *13*. All *AlONSENs* with predicted 4P HREs were upregulated after 6 h HS (Additional file 8). *AlONSEN 3* and *16* were also found upregulated although they did not contain putative HREs. This was most likely caused by ambiguity in RNA-seq analysis, as 100 % of the reads mapping to these elements were multiply mapping to other *ONSENs*. Hence, there is a high correlation between the predicted HREs and RNA-seq results.

Conservation of the most frequent HRE haplotype between the *Arabidopsis* species raised the question about the evolutionary history of *ONSEN* heat-responsiveness in the *Brassicaceae*. Therefore, we searched for *COPIA78* elements in whole-genome assemblies of *Boechera stricta* v1.2, *Brassica rapa* FPsc v1.3 (both JGI; Phytozome), *Capsella rubella* [21], *Eutrema salsugineum* [22], and low coverage assembly of *Ballantinia antipoda* (Vu, Finke, and Pecinka; unpublished data) using genome-wide BLAST searches. We confirmed the absence of *COPIA78* in *Capsella* [23], but found at least one *ONSEN* copy in all other species (Table 1). RT nucleotide sequence identity was >80 % (Additional file 5: Figure S4), fitting previously proposed criteria for a single TE family [24]. The LTR identity was lower (typically <70 %) due to the presence of insertions and deletions and decreased with phylogenetic distance
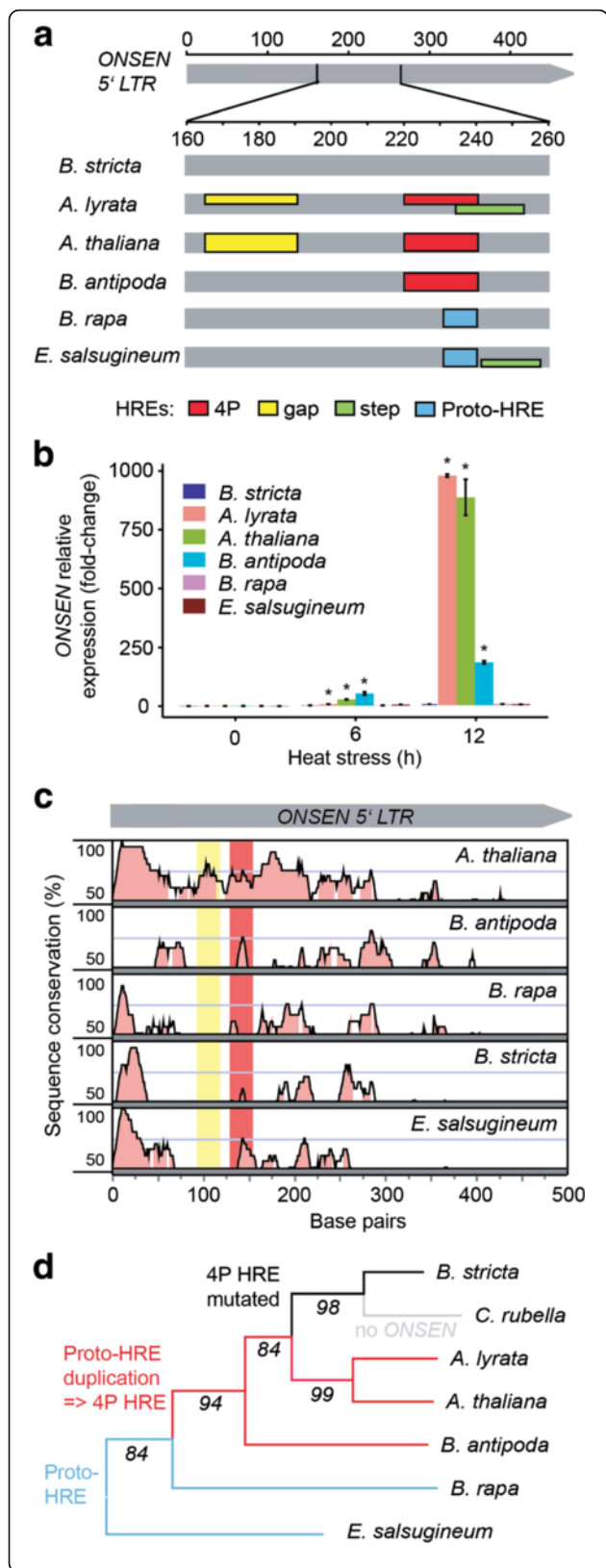
**Table 1** Copy numbers of elements within analyzed *COPIA* families in *Brassicaceae* species

| Species | *ONSEN* | | *COPIA37* | | *HATE* | | *ROMANIAT5* | |
|---|---|---|---|---|---|---|---|---|
| | Total | Full | Total | Full | Total | Full | Total | Full[a] |
| *Arabidopsis lyrata* | 55 | 10 | 57 | 5 | 6 | 6 | 131 | 0 |
| *Arabidopsis thaliana* | 24 | 8 | 32 | 1 | 0 | 0 | 49 | 0 |
| *Ballantinia antipoda* | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Boechera stricta* | 2 | 2 | 0 | 0 | 14 | 7 | 53 | 0 |
| *Brassica rapa* | 6 | 2 | 2 | 0 | 2 | 0 | 7 | 0 |
| *Capsella rubella* | 0 | 0 | 2 | 1 | 0 | 0 | 0[b] | 0 |
| *Eutrema salsugineum* | 2 | 1 | 2 | 0 | 6 | 1 | 65 | 0 |

[a]All *ROMANIAT5* elements lacked integrase domain
[b]Only three solo LTRs were found in *C. rubella*

**Fig. 2** Evolution of *ONSEN* heat-responsiveness. **a** *Schematic representation of in silico reconstruction of putative HREs in ONSEN 5′ LTR in different Brassicaceae species. HRE reconstruction follows criteria proposed by [20]. Colored boxes spanning the entire height of the gray field indicate HREs found in ≥50 % of the heat-responsive copies in A. thaliana and A. lyrata or all copies in other species. The lower boxes represent less frequent (<50 %) variants. Detailed information including sequences underlying individual HREs can be found in Additional file 5: Figure S2.* **b** *Transcript levels of ONSEN elements in Brassicaceae after 6 h and 12 h at 37 °C. Quantitative PCR values were obtained using species-specific primer pairs and normalized to UBC28. Error bars indicate standard deviation of three biological replicates and * P <0.05 in Student's t-test.* **c** *Sequence conservation over the ONSEN 5′ LTR. Species-specific 5′ LTR consensus sequences were compared to A. lyrata query using 20 bp sliding window and 7 bp minimum consensus length. The y-axis for each species shows 50–100 % sequence conservation. Regions with ≥70 % similarity (pink-filled) were considered as conserved. Red and yellow background colors indicate the A. lyrata 4P and gap HRE regions.* **d** *Reconstruction of ONSEN HRE evolution. The phylogenetic tree was developed using the CHALCONE SYNTHASE gene of each individual species. The numbers at branches indicate bootstrap values. Blue lines show species with proto-HREs, red shows those carrying 4P HREs, black shows loss of HREs in B. stricta, and gray shows the COPIA78 family in C. rubella*

(Additional file 5: Figure S5). Nevertheless, it allowed us reconstructing putative HREs. None of the other species contained the *A. thaliana*-like gap HRE (Fig. 2a; Additional file 5: Figure S3). However, there was a perfectly conserved 4P HREs in two out of three *ONSENs* in *B. antipoda*. *ONSENs* of other species either did not contain any HREs (*B. stricta*) or they represented only lower efficiency types and were non-homologous to the *Arabidopsis* HREs (Fig. 2a). To challenge the predicted HREs, we grew all species in vitro and quantified *ONSEN* transcript levels after 6 h and 12 h of HS (Fig. 2b). In agreement with RNA-seq results, we found massive 884–976-fold upregulation in *A. lyrata* and *A. thaliana*. There was also high (185-fold) upregulation in *B. antipoda* containing the putative 4P HRE, but lacking an additional gap HRE (Fig. 2a, b). In contrast, *B. stricta*, *B. rapa*, and *E. salsugineum* predicted to have no or only low efficiency HREs did not show strongly increased *ONSEN* transcript amounts (Fig. 2b).

To test whether HREs represent a major *cis*-regulatory element in *ONSEN* LTRs, we performed phylogenetic shadowing of the LTR consensus sequences (Fig. 2c). Although the 4P HRE region was partially conserved, there are several other similarly conserved regions. The longest stretch of conserved LTR sequence comprises approximately the first 25–30 bp (Fig. 2c), which may be required for TE RT.

By anchoring the structural information on the *Brassicaceae* chalcone synthase-based phylogeny, we reconstructed the evolutionary trajectory of *ONSEN* HREs (Fig. 2d). The nTTCnnGAAn motif, which can be considered as the non-functional sequence preceding the 4P HRE (proto-HRE), is present in *B. rapa* and *E. salsugineum* (Fig. 2a; Additional

file 5: Figure S3), suggesting that it existed already at the onset of *Brassicaceae* evolution. Later, proto-HRE became duplicated and instantly created the high affinity 4P HRE. Molecular dating of the split of the *B. antipoda* lineage [25] suggests that this motif was maintained over 6–9 million years of evolution. However, the 4P HRE was occasionally lost due to accumulation of the point mutations (*B. stricta*) or deletion of whole elements (*Capsella*).

### Species-specific gain of HREs in *COPIA37* and the novel family *TERESTRA*

The other TE family found to be heat-responsive in both *Arabidopsis* species was *COPIA37* (Fig. 1h, i). However, this phenotype was restricted to fewer copies as only 8.8 % (five out of 57) and 12.5 % (four out of 32) of *A. lyrata* and *A. thaliana COPIA37s*, respectively, showed upregulation upon HS (Additional files 8 and 9). The 5′ LTRs of all heat-responsive copies contained putative low affinity binding gap and step HREs (Fig. 3a; Additional file 5: Figure S6). In addition, we found putative 3P HREs in three *AtCOPIA37s* and two *AlCOPIA37s*. These HREs originated from a common nTTCn rich LTR region, but were not identical. Search in other species revealed the presence of *COPIA37* in *B. rapa* (n = 2), *C. rubella* (*n* = 2), and *E. salsugineum* (*n* = 2; Table 1), but here we found only low affinity binding gap and/or step HREs in the latter two species (Fig. 3a; Additional file 5: Figure S6). To test whether the predicted HREs correlate with heat-responsiveness, we exposed all species to 6 and 12 h HS and quantified the transcript amounts by RT-qPCR (Fig. 3b). We observed up to 25-fold activation for *A. lyrata COPIA37* and a weaker (fivefold) activation for *A. thaliana*, both carrying putative 3P HREs. The amount of *COPIA37* transcript reached its peak at 6 h and decreased in spite of continued HS. Other species, carrying only lower efficiency gap and or step HREs, did either not accumulate the transcript or only at a single experimental point. Hence, the most effective 3P HREs evolved independently in *A. lyrata* and *A. thaliana* and also *COPIA37* elements of other species carry diverse set of HREs.

We also identified *TERESTRA* as a new retrotransposon heat-responsive family. The *A. lyrata* genome contains six *TERESTRA* copies sharing 97 % similarity (Fig. 3d). BLAST searches using *TERESTRA* sequences revealed only local similarities to *ONSEN* LTRs and *COPIA46* GAG and POL domains and no other significant hits. Therefore, we performed de novo *TERESTRA* analysis. Based on the order of GAG and POL, *TERESTRA* was unambiguously identified as *Ty1/COPIA* LTR-retrotransposon (Fig. 3c). Based on only 70 % similarity in an alignment of *TERESTRA* to *COPIA46* and *ONSEN* elements (Fig. 3d), we defined *TERESTRA* as a novel *COPIA* family. The consensus length of the complete *AlTERESTRA* element was 5116 bp and the 5′ and the 3′ LTR were 529 and 536 bp long, respectively

(Fig. 3c; Additional file 5: Figure S7). Sequence analyses of *AlTERESTRAs* revealed that all copies are full length, contain a tRNA primer binding site and a polypurine tract, suggesting their autonomy (Fig. 3c). *TERESTRA* LTRs are relatively A-T-rich (69 %) and the consensus sequence contained only a small number of cytosines in symmetrical contexts (CG = 5, CHG = 0; H = A, T or C), which resembles LTR nucleotide composition of *ONSEN* [18]. *TERESTRA* was missing in the *A. thaliana* Col-0 genome. Therefore, we extended our search to 50 *A. thaliana* accessions by genotyping them with *TERESTRA*-specific primers (Additional file 5: Table S3). This screen also gave negative results and suggested the absence of *TERESTRA* in *A. thaliana*. However, we found *TERESTRA* TEs in *Arabidopsis cebennensis* (95 % identity; Additional file 5: Figure S8) and *Arabidopsis halleri* (91 % identity; Additional file 5: Figure S9) using the NCBI sequence database. Furthermore, there were *TERESTRAs* in *B. stricta* (*n* = 14), *B. rapa* (*n* = 2), and *E. salsugineum* (*n* = 6; Table 1), but not outside of the *Brassicaceae*.

All six *AlTERESTRAs* were heat-responsive (Fig. 1h; Additional files 8). Screening of *AlTERESTRA* LTRs for possible HREs revealed a cluster of six nGAAn motifs, which can assemble either two partially overlapping gap HREs or a 4P HRE (Fig. 3e; Additional file 5: Figure S10). Based on the high *AlTERESTRA* transcriptional heat response (Fig. 3f), we favor the latter possibility. Another species with high *TERESTRA* transcriptional activation after 6 h and 12 h HS was *B. stricta* (Fig. 3f). By BLAST we found 14 *TERESTRA* copies in the *B. stricta* genome (Table 1). Eleven copies among them contain a complex cluster of up to five adjacent nGAAn motifs in their 5′ LTRs (Fig. 3d; Additional file 5: Figure S10). According to a conservative approach, three nGAAn motifs within this cluster can putatively form a low affinity gap HRE, but high *TERESTRA* activation in *B. stricta* suggests that all five motifs can establish 4P HREs as compatible with a more relaxed prediction (Fig. 3e, f). Importantly, all predicted *B. stricta* HREs are at positions different from those in *A. lyrata* HREs (Fig. 3e), highlighting their species-specific evolution (Fig. 3g). In *B. rapa*, one *TERESTRA* copy carries a putative gap HRE and another one a step HRE (Fig. 3e). Out of six *TERESTRAs* in *E. salsugineum*, three had putative step and one also an additional gap HRE. However, none of the HREs found in *B. rapa* and *E. salsugineum* was homologous to *Arabidopsis* or *Boechera* HREs and their predicted low HSF binding efficiency was congruent with the absence of heat-responsiveness in the RT-qPCR experiments (Fig. 3f).

Decreased *AlCOPIA37* and *AlTERESTRA* transcript amounts after 12 h versus 6 h HS (Fig. 3b, f) contrasted with continuous transcript accumulation for *AlONSEN* (Fig. 2b). We hypothesized that the failure to maintain high transcript level could be caused by the TGS. Due to
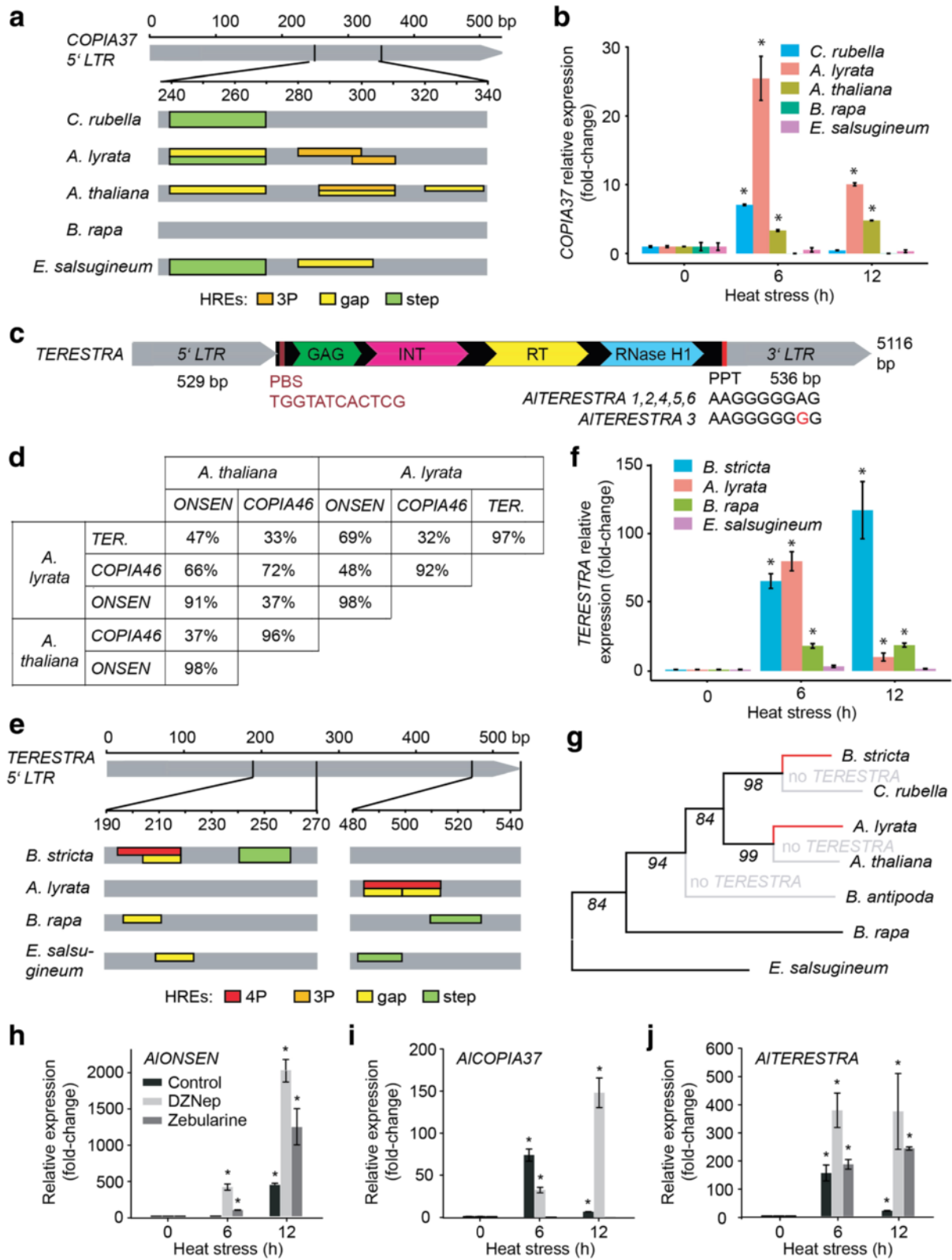
**Fig. 3** (See legend on next page.)

(See figure on previous page.)

**Fig. 3** *COPIA37* and *TERESTRA* are novel heat-responsive *COPIA* families. **a** In silico reconstruction of putative HREs in the 5' LTR of *COPIA37* in different species. HRE classification follows criteria proposed by [20]. *Colored boxes* spanning the entire height of the *gray field* indicate HREs found in ≥50 % of the heat-responsive copies in *A. thaliana* and *A. lyrata* or all copies in other species. The *lower boxes* represent less frequent (<50 %) HREs. Detailed information including sequences underlying individual HREs can be found in Additional file 5: Figure S5. **b** Transcript levels of *COPIA37* in *Brassicaceae* after 6 and 12 h 37 °C HS. The values were normalized to transcript levels of *UBC28*. *Error bars* indicate standard deviation between three biological replicates and * $P$ <0.05 in Student's t-test. **c** *Schematic representation* of *A. lyrata TERESTRA* (*TERESTRA*). LTRs are indicated in gray. Capsid protein (GAG), integrase (INT), RT, and RNAse H1 domains are shown within the *light-blue-labeled TERESTRA* protein-coding part. Primer binding sequence (PBS) and polypurine tract (PPT) are indicated by *red boxes*. **d** Sequence similarities within pair-wise LTR alignments between *A. lyrata* and *A. thaliana TERESTRA*, *ONSEN*, and *COPIA46* families. More than 70 % similarity was expected for members of the same family. *TERESTRA* is absent in *A. thaliana*. **e** In silico reconstruction of putative HREs in the 5' LTR of *TERESTRA*. The criteria were as described for Fig. 3a. Detailed information including sequences underlying individual HREs can be found in Additional file 5: Figure S9. **f** Transcript levels of *TERESTRA* in response to 6 and 12 h 37 °C HS in different *Brassicaceae*. The experiment was performed as described in (**b**). **g** Reconstruction of *TERESTRA* HRE evolution. The phylogenetic tree was developed using a chalcone synthase gene of each individual species. The numbers at the base of the branches indicate bootstrap values. *Black lines* show species with low efficiency HREs and *red lines* highlight independently evolved high efficiency HREs in *A. lyrata* and *B. stricta*. *Gray lines* denote species where *TERESTRA* could not be found. TE transcript accumulation of (**h**) *ONSEN*, (**i**) *COPIA37*, and (**j**) *TERESTRA* after 0, 6, and 12 h 37 °C HS preceded by 48 h control (no inhibitor), 10 μM 3-deazaneplanocin A (DZNep), or 40 μM zebularine treatment. Transcript amounts were normalized to *UBC28* mRNA and signals from drug and heat-treated samples were recalculated as fold-changes relative to 0 h. *Error bars* indicate variation between two biological replicates and * $P$ <0.05 in Student's t-test

a lack of *A. lyrata* TGS mutants, we used a pharmacological approach to interfere with TE silencing [26]. We treated 14-day-old *A. lyrata* plants with 10 μM 3-deazaneplanocin A (DZNep) and 40 μM zebularine, including control plants without treatment. DZNep is an S-adenosylhomocysteine synthesis inhibitor, which blocks the production of SAM, the methyl group donor required for DNA and histone methylation. Zebularine is a cytidine analog leading to DNA de-methylation and loss of silencing from specific transposons [27–29]. After two days of drug treatment, plants were heat-stressed for 0, 6, and 12 h and the amount of transcript analyzed by RT-qPCR (Fig. 3h–j). Control DZNep and zebularine treatment increased *COPIA37* and *ONSEN* transcript tenfold and fivefold, respectively (Additional file 5: Figure S11), suggesting that both TEs can be weakly activated by TGS interference also without HS treatment. *TERESTRA* was not activated by the drug treatment. A combination of HS with drug treatments had strong additive effects in all cases, except for zebularine and HS-treated *COPIA37* (Fig. 3h–j). Both *ONSEN* and *TERESTRA* transcripts accumulated at much higher levels that were not decreasing at 12 h HS (Fig. 3h, i). The effect was generally stronger for DZNep and weaker for zebularine. This suggests that the heat-induced TE transcript accumulation is rapidly suppressed by epigenetic means, in particular for TEs carrying lower affinity binding HREs.

## AtROMANIAT5 contributes to transcriptional regulation of APUM9 under HS

There are four heat-responsive *ROMANIAT5* TEs in *A. thaliana* but none in *A. lyrata* (Fig. 1h, i; Additional files 8 and 9). All *AtROMANIAT5* elements lack an integrase domain, suggesting that these elements are incomplete and non-autonomous (Table 1). A previous study revealed that one of the heat-responsive copies *AtROMANIAT5-*

2 (At1g35735) is under complex epigenetic control by Morpheus molecule 1 (*MOM1*) and RdDM pathways, and loss of this control causes upregulation of the *Arabidopsis PUMILIO9* (*APUM9*; At1g35730) gene located directly downstream of the TE [30]. To better understand the potential role of *ROMANIAT5* in regulating *APUM9* during HS, we reconstructed their loci in *A. thaliana* and *A. lyrata* (Fig. 4a, b) and also retrieved the number of reads mapping to both loci under different experimental conditions (Fig. 4c, d). Interestingly, we observed significant (t-test, $P$ <0.05) upregulation of *APUM9* upon HS in *A. thaliana* but not in *A. lyrata* (Fig. 4c, d), where the nearby *ROMANIAT5* is missing. This suggested that *ROMANIAT5-2* controls *APUM9* transcription under HS. To validate this observation, we used a reporter line (called Silex) which contains the *APUM9* upstream region and the *ROMANIAT5-2* 3' LTR upstream of a GFP reporter (Fig. 4a) [31]. The Silex reporter construct is silenced during entire *A. thaliana* development, except for developed siliques, but the reporter activity can be restored in the background of *MOM1* RdDM double mutants and histone deacetylase 6 mutants [31]. We grew Silex reporter plants under controlled conditions with and without HS. GFP transcripts were missing in the control plants but present after 12 and 24 h at 37 °C (Fig. 4e). GFP accumulated in the apical meristem after 24 h of HS recovery and remained detectable for at least five days (Fig. 4f, g), although GFP transcript was not present anymore (Fig. 4e). Heat-responsiveness of Silex transgene in the absence of *ROMANIAT5* 5' LTR suggested that the locus may be at least partially controlled by a bi-directional heat-responsive promoter activity of the 3' LTR. Indeed, we found putative 3P/gap HREs within the 3' LTRs (and also the 5' LTRs) of all heat-responsive *AtROMANIAT5* TEs. However, transcription from the 3' LTR could result in *ROMANIAT5-2* antisense transcript. To
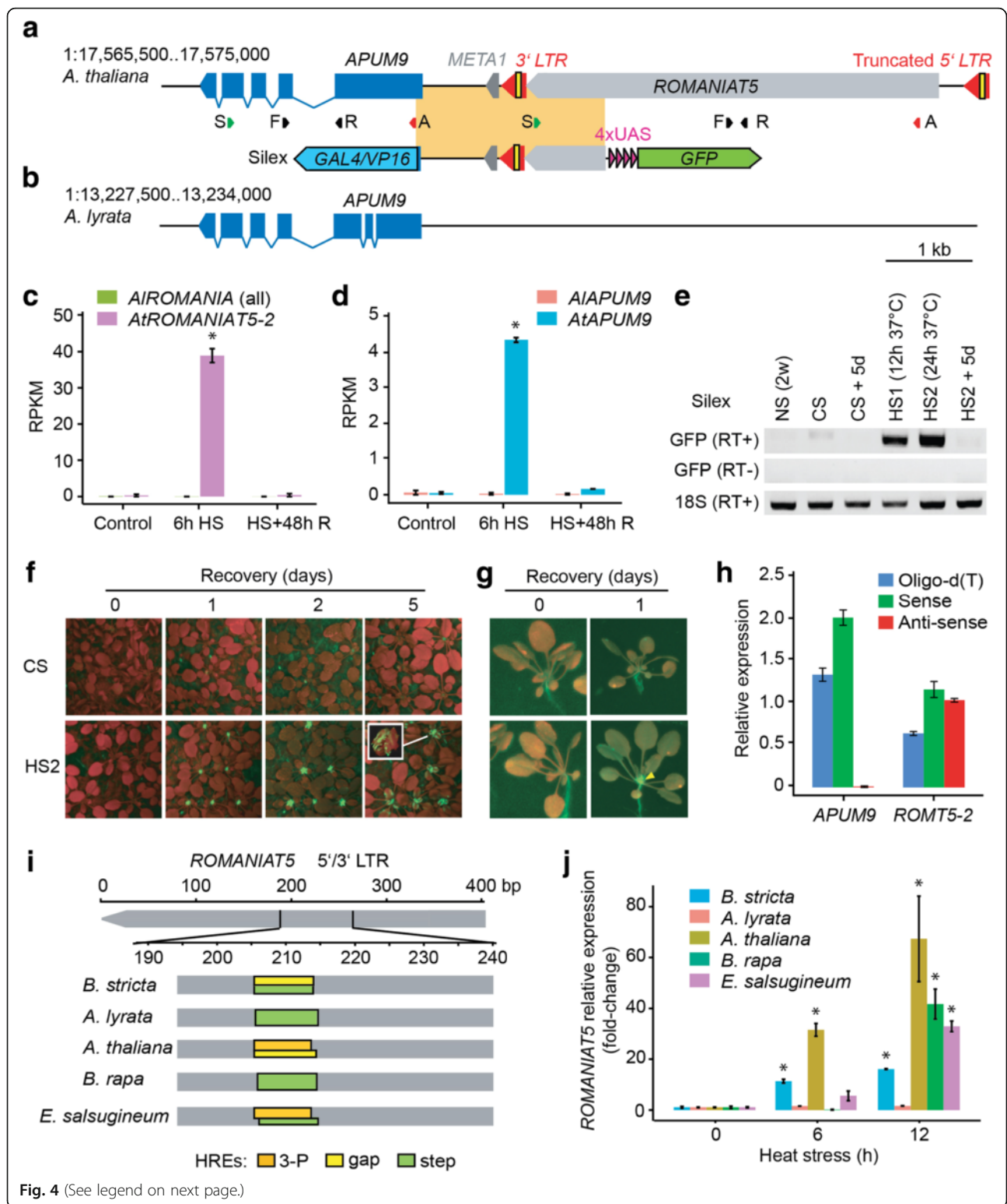
**Fig. 4** (See legend on next page.)

(See figure on previous page.)
**Fig. 4** *ROMANIAT5-2* controls heat-responsiveness of *APUM9* in *A. thaliana*. **a** *Schematic representation* of the *ROMANIAT5-2 – APUM9* region in *A. thaliana*. The *yellow block* within the 3' LTR represents a 3P/Gap heat responsive element (HRE). *S* position of primers for RT of the sense transcripts, *A* position of primers for RT of the anti-sense transcripts, *F* and *R* forward and reverse quantitative PCR primers. *META1* is a transposon fragment flanking *ROMANIAT5-2* 3' LTR. Silex: the *orange block* corresponds to the genomic fragment cloned upstream of the 4× "upstream activating sequence" (UAS, *violet*) and green fluorescent protein (*GFP*; *green*). **b** *Schematic representation* of the *A. lyrata APUM9* locus. Reads per kilobase per million reads (RPKM) for (**c**) *ROMANIAT5* and (**d**) *APUM9* under control, 6 h at 37 °C HS and HS with 48 h recovery at control conditions (HS + R). * *P* <0.05 in t-test. **e** RT-PCR analysis of Silex reporter construct response to HS. *NS* non-stressed control plants, *CS* and *HS* control- or heat-stressed plants, respectively, *+0* and *+5d* days of recovery at non-stress conditions, *RT+* and *RT–* samples with and without RT, respectively. 18S rRNA transcript serves as positive control. **f** GFP signal in control and 24 h heat-stressed (HS2) Silex, detected after 0, 1, 2, or 5 days of recovery. *Red* – chlorophyll, *green* – GFP. **g** *Close-up view* of plants treated as described in (**f**). **h** Strand-specific RT-qPCR of *APUM9* and *ROMANIAT5-2* in *A. thaliana* after 6 h HS. **i** Putative HREs in *ROMANIAT5* LTRs in *Brassicaceae*. **j** RT-qPCR for *ROMANIAT5* in *Brassicaceae* after 6 and 12 h at 37 °C HS. The values were normalized to *UBC28*. *Error bars* indicate standard deviation between three biological replicates and * *P* <0.05 in Student's t-test

test this, we isolated *A. thaliana* RNA after HS and performed complementary DNA (cDNA) synthesis using strand-specific RT primers (Fig. 4a). Control cDNA from RT with oligo-d(T) primers gave signals for both genetic elements (Fig. 4h). Strand-specific RT-qPCR revealed HS-induced sense transcript, but no antisense transcript, for *APUM9*. In contrast, both types of primers resulted in amplification of *ROMANIAT5-2* transcripts, suggesting that it is transcribed in both directions under HS. The distribution of RNA-seq reads did not indicate large amounts of a read through transcription from *ROMANIAT5-2* to *APUM9* (Additional file 5: Figure S12). Altogether, this confirms the 3' LTR as bi-directional HS-responsive promoter.

We found *ROMANIAT5* elements in genomes of all species except for *B. antipoda* (Table 1). Putative HREs were present in at least some copies of *ROMANIAT5* in all species except for *C. rubella* that contained only solo LTRs. There were step and gap HREs in *A. lyrata*, *B. rapa*, and *B. stricta*, 3P/gap HREs in *E. salsugineum*, and 3P HREs in *A. thaliana* (Fig. 4i; Additional file 5: Figure S13). The predicted HSF binding affinity of individual HREs correlated well with the amount of *ROMANIAT5* transcripts found after 6 and 12 h of HS (Fig. 4j). The only exception was *B. rapa*, which showed 42-fold upregulation after 12 h HS but the analyzed copies carried at most only low affinity step HRE. This could be due to the presence of heat-responsive *ROMANIAT5* copies in the part of the *B. rapa* genome that is not yet assembled.

## Discussion

Transpositions and insertions of TEs may lead to loss of gene functionality [32, 33]. Therefore, TEs activity and mobility are tightly controlled by epigenetic means throughout the entire plant development [5, 6]. On the other hand, new insertions contribute to genome evolution and regulation of gene transcription [2]. Therefore, it was already suggested in the early days of transposon research that, under conditions when diversity of regulatory patterns in a population may provide a better basis for selection, limited transposon activation could be beneficial [34]. However, how occasional TE expression is provoked and how

control is regained later is still a matter of debate. There is a rapidly increasing number of reports showing transient TE activation under various stress conditions reviewed in [1]. Therefore, it was hypothesized that stresses may open a window for transpositions. Here, we introduced the *ONSEN* (*COPIA78*) family as a model for understanding TE control and behavior under HS. *ONSEN* shows massive transcriptional upregulation upon HS in *A. thaliana* and new insertions in progenies of heat-stressed Pol IV mutant [7, 8, 17]. The molecular basis of *ONSEN* heat-responsiveness was puzzling until recently, when a typical HSFA2 TF binding HRE was identified in its *cis*-regulatory region [18]. Presence of canonical TF binding motifs in TE promoters was described for *D. melanogaster* and *M. truncatula* [15, 35]. However, the frequency of such activation strategy among TEs was unknown.

We analyzed LTRs of *A. thaliana* and *A. lyrata* heat-responsive *COPIA* TEs *ONSEN*, *COPIA37*, *TERESTRA*, and *ROMANIAT5* for putative HREs. A minimum of three adjacent (<5 bp) nGAAn motifs can form a basal HRE, whose activity will depend on their distance and the total number [20]. Heat-responsive *COPIAs* featured the whole spectrum of HREs ranging from the 4P types in *ONSEN* and *TERESTRA*, through 3P types in *COPIA37* and *ROMANIAT5* to a dozen of variable gap and step HREs in all these families. By comparing predicted HREs with transcriptional data, we conclude that gap and step HREs are mostly not sufficient to trigger HS-induced TE upregulation. This is congruent with their proposed low HSF binding efficiency [36]. Predicted 3P HREs correlated with up to a hundred-fold (*COPIA37*, *ROMANIAT5*) and 4P HREs with up to a thousand-fold (*ONSEN*, *TERESTRA*) transcript accumulation upon HS. This suggests a strong correlation between putative HREs and the transcriptional response of the TEs.

Previously it was shown that the TGS machinery antagonizes the TE activation [7, 17]. We found that the speed of re-silencing during or after HS depends on the HRE type. While *ONSEN*, with the strong 4P HRE, accumulated transcript during entire HS exposure, TEs carrying lower affinity HREs typically showed a maximum transcript amount at 6 h HS and lower levels at 12 h

HS. This silencing can be reduced by treatment with DNA methylation inhibitors. Hence, stressed plants take active measures to prevent TE transpositions already during ongoing HS treatment. However, HS-induced TE activation must not always aim at transposition, but can be part of the plant regulome [2]. In *A. thaliana*, we found that heat-responsive *AtROMANIAT5-2* controlled transcription of the *APUM9* gene located downstream of the element. As we did not observe any evidence for high amount of a read-through transcript from *ROMANIAT5-2* towards *APUM9*, we hypothesize that this transcriptional activation may be mediated rather by a specific three-dimensional chromatin organization at this locus. *APUM9* gene was previously shown to be under control by *HDA6* and synergistically by *MOM1* and RdDM pathways, but not *DDM1* and *MET1* [30, 31]. Therefore, *AtROMANIAT5-2* may represent a domesticated transposon with fine-tuned HS-regulated activation, contributing to transcriptional control of *APUM9*.

To challenge the hypothesis that HREs could be beneficial for TE amplification (but not necessarily for the host genome stability), we reconstructed evolutionary trajectories for HREs of *ONSEN, COPIA37, TERESTRA,* and *ROMANIAT5* in the *Brassicaceae. ONSEN* was not heat-responsive in the early separated lineages represented by *B. rapa* and *E. salsugineum*, because its LTRs contained only one half of the 4P HRE (proto-HRE), which does not constitute a functional HRE. The proto-HRE became duplicated approximately 6–9 millions of years ago [25] and directly formed the present days 4P HRE found in the genus *Arabidopsis* and in the Australian species *B. antipoda*. Hence, *ONSEN* 4P HRE represents an evolutionary conserved *cis*-regulatory element. However, it should be noted that there are several other similarly or even more conserved regions within the *ONSEN* LTR. Whether they represent other TF binding sites and/or enhancers remains currently unknown. Furthermore, the *ONSEN* example shows that even high affinity HREs do not allow a TE to overrule the host genome defense, because their heat-responsiveness was lost in *B. stricta*, and the whole family became vanished from the *C. rubella* genome. In *TERESTRA,* high affinity 4P HREs evolved independently at two different LTR regions in the closely related species *A. lyrata* and *B. stricta*, while 3P HREs of *COPIA37* emerged multiple times from a common nTTCn-rich LTR region. In contrast to *ONSEN*, HREs of these families are evolutionary young and species-specific. Whether they will be evolutionary successful, is an open question, but we speculate this to be the case for *A. lyrata TERESTRA*, where all genomic copies are full length, carry strong HRE, and respond to heat.

At present it is unknown whether higher temperatures in southern latitudes lead to greater amplification of heat-responsive TEs in subtropical relative to temperate zones. Although this is possible, there are also several factors that may act against such correlation. First, southern populations may reduce effects of HS by adaptation and growth at favorable microclimatic and/or temporal conditions [37]. Second, the genomes are subject to purification mechanisms and the higher transposition rate may be opposed by a greater frequency of TE removal [10]. Indeed, HS was shown to increase frequency of DNA sequence removal by a single strand annealing type of homologous recombination in transgenic constructs structurally resembling a LTR retrotransposon [38, 39]. Therefore, the final number of stress responsive TEs per genome may be the result of multiple effects acting in a complex network.

## Conclusions

TEs evolve *cis*-regulatory elements, such as HREs, rapidly and independently in many groups. This may represent a strategy to produce new copies, constantly challenging the host defense system by searching for potential weak points. Successful regulatory elements may become evolutionary conserved and spread by new TE insertions in a self-reinforcing loop. However, these copies will be silenced and frequently removed from the genome. Hence, stress-mediated TE activation is likely not an unequivocal and straightforward winning principle, but rather a necessary strategy to survive under the pressure of the host defense systems. It is also likely that the host genome can benefit to some extent, and in specific cases, from *cis*-regulatory elements spread by TEs.

## Methods

### Plant materials and growth conditions

We used: *Arabidopsis thaliana* Col-0 and Silex [31], *Arabidopsis lyrata* subsp. *lyrata* MN47, *Ballantinia antipoda*, *Boechera stricta* ES9, *Brassica rapa* FPSc, *Capsella rubella*, and *Eutrema salsugineum*. Before standard HS experiments, *A. thaliana* and *A. lyrata* seeds were placed on wet soil, stratified for one week at 4 °C, and then grown in a growth chamber (Percival) at 21 °C during the day and 16 °C during the night (16 h light/8 h dark) until plants reached approximately the five-leaves stage. Subsequently, a part of the plants was transferred to 37 °C for 6 h. RNA samples for sequencing were collected from some of the stressed plants and the controls directly after stress. The remaining stressed plants were allowed to recover at control conditions and collected after 48 h. Later, HS and drug-treatment experiments were performed with in vitro grown plants. First, the seeds were surface-sterilized with 8 % sodium hypochlorite for 6 to 12 min, washed with copious amounts of sterile water, dried under sterile conditions, and spread on sterile ½ Murashige-Skoog medium. After one week of stratification at 4 °C, the Petri dishes with seeds were transferred to growth chamber with a long

day regime (16 h light/8 h dark) and constant temperature of 21 °C. Plates with rosettes at the pre-bolting stage were then placed in another chamber with 37 °C for 6 h. For combined drug and heat treatments, *A. lyrata* plants were grown as described above, then transferred to plates with no inhibitor, 10 μM DZNep, or 40 μM zebularine (both Sigma-Aldrich) for 48 h and then exposed to 0, 6, or 12 h at 37 °C HS. Aerial plant tissues were harvested immediately after the stress, flash frozen in liquid nitrogen, and stored at −80 °C.

Seeds of the Silex reporter line were sown directly on potting soil and stratified at 4 °C for 48 h. The pots were then placed in a Percival CU-22 L chamber at 21 °C with 12 h light (140 mmol m$^{-2}$s$^{-1}$) and 12 h dark. When the plants turned 14 days old, the pots were placed at 6 °C under the same light conditions for 24 h. At this time, control plants were moved back to the 21 °C chamber while HS plants underwent 24 h HS at 37 °C with light conditions as before. Immediately after the HS treatment, all pots were placed again at 21 °C. Fluorescence pictures of control and HS plants were taken at 0 and after 1, 2, and 5 days of recovery. Fluorescence imaging was performed using an Aequoria dark box with a mounted ORCAII CCD camera (Hamamatsu, Japan).

### Nucleic acids extraction, cDNA synthesis, and RT-qPCR

Genomic DNA was isolated using the Phytopure gDNA Kit (GE Healthcare). Total RNA was isolated with the RNeasy Plant Mini Kit (Qiagen) with an on column DNaseI (Roche) digestion or by the standard Trizol method with additional DNaseI (Thermo Scientific) digestion. cDNA was synthesized from 1 μg total RNA per sample using the Revert Aid H-Minus First Strand cDNA synthesis kit with the oligo-d(T) primer (all Thermo Scientific). For strand-specific RT, total RNA of 6 h HS *A. thaliana* plants was divided into five aliquots which were converted into cDNA using (1) oligo-d(T) primer, *APUM9* (2) sense and (3) antisense transcript primer, and *ROMANIAT5-2* (4) sense and (5) antisense transcript primers. RT-qPCR analysis was performed on three biological replicates with at least two technical replicates in a CFX384 instrument (BIO-RAD) using the SensiMix Plus SyBr Kit (PEQLAB). Expression values were calculated relative to control-treated samples using the standard curve method [40] and normalized using the glyceraldehyde-3-phosphate-dehydrogenase C2 (*GAPC-2*) or the *UBC28* gene with a stable expression under mock, HS, and recovery conditions. Primers used in this study are listed in Additional file 5: Table S4.

### RNA sequencing

One μg total RNA per sample with RIN >8.0 (Agilent Bioanalyzer 2100) was used to construct strand non-specific sequencing libraries with the Illumina TruSeq RNA Library Kit v2 according to the manufacturer's

instructions. Library quality was tested on a Bioanalyzer and high-quality libraries were subsequently sequenced in the 100 bp single-end read mode using a HiSeq 2500 sequencer (Illumina). Adaptor sequences and low quality bases were trimmed and low quality reads were filtered out with the FAST-X toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) using custom-made scripts. Subsequently, reads were mapped to the corresponding reference genome (TAIR 10 genome assembly or *A. lyrata* genome assembly v1.0) using tophat2 [41] with default settings. On average, >15 million sequencing reads per library passed trimming and quality filtering. The numbers of reads mapping to specific genomic positions were retrieved using Qualimap and the latest *A. thaliana* genome annotation TAIR10 and *A. lyrata* genome annotation v2 [42] for genes and custom-made repeat annotations for TEs. The TE data were further processed with COMEX (see below) and data for genes were analyzed directly using the DESeq package in R software [43, 44].

### COMparative EXpression of transposable elements (COMEX)

Accurate quantification of TEs expression using short read sequences is hampered by high similarity of potentially many genomic copies. We developed a simple protocol called COMEX (https://github.com/bpietzenuk/COMEX) that partially overcomes this problem and allows analysis of TE transcription from RNA-seq data. Out of >10 million reads per average sequencing library, 0.12 % and 0.73 % high-quality mappable reads corresponded to TEs within our custom made *A. thaliana* and *A. lyrata*, respectively, TE annotations. This suggests that TE expression analysis using RNA-seq can be made more sensitive by high sequencing depth. The reads were processed via a shell-script that merges the pipeline as follows. First, the binary mapping.bam file is converted into a readable .sam file. Subsequently, ends are printed (ToPrint_end1.py) to the .sam file and mapping errors are removed (Selectnonrepeated1.py). In the following step (Selectmultiplymapped1.py), the output files for the uniquely mapping and the multiply mapping reads are created. The high-quality uniquely mapping (UM) TE reads were directly accepted for expression analysis. High-quality multiply mapping TE reads were analyzed to identify those providing usable information. We classified multi-mapping reads into two categories: (1) informative reads mapping to multiple members of the same TE family (Specifically Multiply Mapped – SMM); and (2) non-informative reads mapping across TE families (Non-specifically Multiply-Mapped – NMM) using the TE annotation gff-file. Reads of the second category were discarded (new_cases1.py). Afterwards, UM and SMM are merged into a single .sam-file and converted into a binary .bam-file. Subsequently, the output file of the COMEX2.0-pipeline (filename.output.final.bam) containing the number of SMM and UM reads from the

same TE family was retrieved using a strand non-specific protocol in Qualimap. To avoid a bias by repeated counting of SMM reads, we used the proportional read count method that divides the power of a read by the number of mapped positions. This provided the number of reads per individual TE families and TEs, which were subjected to statistical analysis using the DESeq package in R software [43]. To avoid considering potentially large number TEs with minimal transcriptional changes, which would be later difficult to validate experimentally, we considered only those which had at least 0.55 RPKM in one of the experimental time points.

### In silico sequence analysis

Sequences of interest were extracted from corresponding TE annotation files using bedtools [45]. LTR reconstruction was carried out in LTR-Finder [46] or manually by pairwise and multiple alignments of the 3'end to the 5'ends of TE annotated regions using MUSCLE or multalin with the DNA 5–0 comparison table option. Structural analysis and annotation of *TERESTRA* was performed using LTR Finder and blastx using NCBI non-redundant protein sequences database. LTR_Finder was used in both analyses with the threshold option set to 2.0 using the tRNA database of *A. thaliana* to predict PBS. The LTR length range was set from 100–3500 and the minimum LTR distance was set to 1000. Other parameters were left at default settings. Search for *ONSEN* sequences within genomes of various *Brassicaceae* was done using BLASTN within Phytozome 10 [47, 48]. Hits with a sequence identity of >70 % were extracted and manually investigated. Positive hits with a query coverage <70 % were analyzed manually for sequence similarity with Multialign using the DNA 5–0 comparison table option. The input ONSEN RT and LTR sequences are provided in Additional files 10 and 11, respectively.

### Phylogenetic analysis

To analyze the evolutionary distance of the *Ty1/COPIA* LTR-retroelements, multiple sequence alignments of the RT domains were performed using the genomic nucleotide sequences in MUSCLE [49]. RT protein sequences used for construction of the network and the tree (Fig. 1j and Additional file 5: Figure S2, respectively) are provided in Additional file 12. The evolutionary history was inferred using the Neighbor-Joining method (Kimura-2-Parameter method) with 1000 bootstrap replicates. Positions containing missing data and gaps were removed (pairwise deletion option) leading to a total of 862 position in the final dataset. The tree was visualized as an unrooted tree. Phylogenetic network of genomic RT domain sequences from *Ty1/COPIA* LTR-retroelements was constructed using Neighbor-Net [50] within the splitstree 4.0 package [51, 52]. The phylogenetic

distances were calculated by LogDet-pairwise genetic distances using LDDIST [53] with imputed missing matrix entries. Multiple sequence alignments of *CHS* genomic sequences were performed using MUSCLE [49]. The *CHS* phylogeny was inferred using the Maximum Likelihood tree based on the Kimura-2-parameter model with 1000 bootstrap replicates. *CHS* sequences were retrieved from [25]. All positions containing gaps and missing data were eliminated. There were a total of 1267 positions in the final dataset. All phylogenetic trees were constructed within MEGA 7 [54]. Phylogenetic shadowing and analysis of motif conservation was performed with mVISTA [55, 56] using LTR consensus sequences of different species prepared in BioEdit [57], allowing fasta ambiguity codes for low conserved positions. Sequences were aligned using AVID [58]. The cutoff was defined as ≥70 % conservation over a 20 bp sliding window with the minimal consensus of 7 bp relative to *A. lyrata* 5' LTR sequence.

### Accession numbers

Short sequence reads were deposited in the NCBI GEO archive under accession number GSE69077.

### Additional files

**Additional file 1:** Table listing transcriptional changes of genes after 6 h HS and recovery in *A. lyrata*. (XLSX 6774 kb)

**Additional file 2:** Table listing transcriptional changes of genes after 6 h HS and recovery in *A. thaliana*. (XLSX 6407 kb)

**Additional file 3:** List of *A. lyrata* TEs in general feature format. (GFF 4554 kb)

**Additional file 4:** List of *A. thaliana* TEs in general feature format. (GFF 1310 kb)

**Additional file 5: Figure S1.** Number of TE families in *A. thaliana* (n = 364) and *A. lyrata* (n = 376) as identified by RepeatMasker. **Figure S2.** Unrooted phylogenetic tree of heat-responsive and -non-responsive *COPIA* TEs. **Table S1.** List of *ONSEN* elements in *A. thaliana* Col-0 genome. **Figure S3.** HREs found in 5' LTRs of *ONSEN* elements. **Table S2.** List of *COPIA78/ONSEN* elements in *A. lyrata* MN47 genome. **Figure S4.** Percentage identity matrix of the RT nucleotide sequences of *ONSEN* elements from different species. **Figure S5.** Percentage identity matrix of LTR nucleotide sequences of *ONSEN* elements from different species. **Figure S6.** HREs found in the 5' LTRs of *COPIA37* elements. **Figure S7.** Consensus DNA sequence of *A. lyrata* *TERESTRA*s. **Table S3.** List of *A. thaliana* accessions negatively tested for presence of *TERESTRA* elements. **Figure S8.** The fragment of *TERESTRA* from *A. cebennensis* clone 44. **Figure S9.** *A. halleri* *TERESTRA* reconstructed based on NCBI BLASTs using *A. lyrata* *TERESTRA* consensus sequence. **Figure S10.** HREs found in 5' LTRs of *TERESTRA* elements. **Figure S11.** Transcriptional response of *ONSEN*, *COPIA37*, and *TERESTRA* to DNA methylation inhibitor treatments in *A. lyrata*. **Figure S12.** Density of RNA-seq reads mapping over *APUM9* – *ROMANIAT5-2* region. **Figure S13.** Putative HREs in 5'/3' LTRs of *ROMANIAT5* elements. (PDF 2864 kb)

**Additional file 6:** Table listing transcriptional changes of all TEs after 6 h HS and recovery in *A. lyrata*. (XLSX 6086 kb)

**Additional file 7:** Table with transcriptional changes of all TEs after 6 h HS and recovery in *A. thaliana*. (XLSX 1850 kb)

**Additional file 8:** Table showing significantly upregulated and downregulated TEs after 6 h HS and recovery in *A. lyrata*. (XLSX 63 kb)

**Additional file 9:** Table listing significantly upregulated and downregulated TEs after 6 h HS and recovery in *A. thaliana*. (XLSX 825 kb)

**Additional file 10:** Contains RT sequences of *ONSENs* from different species and was used to generate Additional file 5: Figure S4. (FASTA 15 kb)

**Additional file 11:** Includes LTR sequences of *ONSENs* from different species and was used to generate Additional file 5: Figure S5. (FASTA 14 kb)

**Additional file 12:** Contains RT amino acid sequences of *COPIA* TEs used for to generate Fig. 1j and Additional file 5: Figure S2. (FASTA 73 kb)

### Availability of data and materials
RNA-sequencing data have been uploaded to the Gene Expression Omnibus and are available under accession numbers [GEO: GSE69077]. Lists of *A. lyrata* and *A. thaliana* TEs are provided in Additional files 3 and 4. Primers for PCR analyses are available in Additional file 5: Table S4. The COMEX method is described in detail and is available to download from https://github.com/bpietzenuk/COMEX; this software is licensed under GNU General Public License version 3. The seeds of plants used in this work can be requested from the corresponding author.

### Authors' contributions
Experiments in this study were conceived by AP, EB, and AMJ. Experiments were performed by BP, CM, HG, NB, and AA. RNA sequencing was analyzed by BP and NB. Phylogenetic analyses were performed by BP. Transcript quantifications were done by CM, BP, and HG. Silex experiments were performed by HG. AP wrote the manuscript with contribution of all authors. All authors read and approved the final manuscript.

### Competing interests
The authors declare that they have no competing interests.

### Ethics approval and consent to participate
Not applicable.

### Author details
[1]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne 50829, Germany. [2]Present address: Department of Plant Physiology, Ruhr-University Bochum, Bochum, Germany. [3]Department of Crop Science, Federal University of Rio Grande do Sul, Porto Alegre, RS 91540000, Brazil. [4]Department of Plant Biology, University of Geneva, Sciences III, 30 Quai Ernest-Ansermet, 1211 Geneva 4, Switzerland. [5]Present address: The Sainsbury Laboratory, University of Cambridge, Cambridge, UK. [6]Present address: Cardiovascular proteomics, Centro Nacional de Investigaciones Cardiovasculares, Madrid 28029, Spain. [7]UMR1345 IRHS, Université d'Angers, INRA, Université Bretagne Loire, SFR4207 QUASAV, 49045 Angers, France.

### References
1. Grandbastien M-A. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2015;1849:403–16.
2. Lisch D. How important are transposons for plant evolution? Nat Rev Genet. 2013;14:49–61.
3. Vu GTH, Schmutzer T, Bull F, Cao HX, Fuchs J, Tran TD, et al. Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. Plant Genome. 2015;8:0021. doi:10.3835/plantgenome2015.04.0021.
4. Mari-Ordonez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. Reconstructing *de novo* silencing of an active plant retrotransposon. Nat Genet. 2013;45:1029–39.
5. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11:204–20.
6. Matzke MA, Mosher RA. Rna-directed DNA methylation: An epigenetic pathway of increasing complexity. Nat Rev Genet. 2014;15:394–408.
7. Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Scheid OM. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell. 2010;22:3118–29.
8. Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I. Stress-induced activation of heterochromatic transcription. PLoS Genet. 2010;6:e1001175.
9. Baubec T, Finke A, Mittelsten Scheid O, Pecinka A. Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. EMBO Rep. 2014;15:446–52.
10. Devos KM, Brown JKM, Bennetzen JL. Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res. 2002;12:1075–9.
11. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet. 2011;43:476–81.
12. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. Evolution of DNA methylation patterns in the *Brassicaceae* is driven by differences in genome organization. PLoS Genet. 2014;10:e1004785.
13. Willing E-M, Rawat V, Mandáková T, Maumus F, James GV, Nordström KJV, et al. Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. Nature Plants. 2015;1:14023.
14. Gutzat R, Mittelsten Scheid O. Epigenetic responses to stress: Triple defense? Current Opin Plant Biol. 2012;15:568–73.
15. Ivashuta S, Naumkina M, Gau M, Uchiyama K, Isobe S, Mizukami Y, et al. Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (*Medicago sativa*). Plant J. 2002;31:615–27.
16. Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol. 2009;149:88–95.
17. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature. 2011;472:115–9.
18. Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O. How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genet. 2014;10:e1004115.
19. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153:193–205.
20. Sakurai H, Enoki Y. Novel aspects of heat shock factors: DNA recognition, chromatin modulation and gene expression. FEBS J. 2010;277:4140–9.
21. Slotte T, Hazzouri KM, Agren JA, Koenig D, Maumus F, Guo Y-L, et al. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. Nat Genet. 2013;45:831–5.
22. Yang R, Jarvis DJ, Chen H, Beilstein M, Grimwood J, Jenkins J, et al. The reference genome of the halophytic plant *Eutrema salsugineum*. Frontiers Plant Sci. 2013;4:46.
23. Ito H, Yoshida T, Tsukahara S, Kawabe A. Evolution of the *ONSEN* retrotransposon family activated upon heat stress in *Brassicaceae*. Gene. 2013;518:256–61.
24. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.
25. Mandakova T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA. Fast diploidization in close mesopolyploid relatives of Arabidopsis. Plant Cell. 2010;22:2277–90.
26. Pecinka A, Liu C-H. Drugs for plant chromosome and chromatin research. Cytogenet Genome Res. 2014;143:51–9.
27. Baubec T, Pecinka A, Rozhon W, Mittelsten SO. Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. Plant J. 2009;57:542–54.
28. Foerster AM, Dinh HQ, Sedman L, Wohlrab B, Mittelsten SO. Genetic rearrangements can modify chromatin features at epialleles. PLoS Genet. 2011;7:e1002331.
29. Liu C-H, Finke A, Diaz M, Rozhon W, Poppenberger B, Baubec T, et al. Repair of DNA damage induced by the cytidine analog zebularine requires ATR and ATM in Arabidopsis. Plant Cell. 2015;27:1788–800.

30. Yokthongwattana C, Bucher E, Čaikovski M, Vaillant I, Nicolet J, Scheid OM, et al. MOM1 and POL-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. EMBO J. 2010;29:340–51.

31. Hristova E, Fal K, Klemme L, Windels D, Bucher E. HISTONE DEACETYLASE 6 controls gene expression patterning and DNA methylation-independent euchromatic silencing. Plant Physiol. 2015;168:1298–308.

32. Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, et al. Selective epigenetic control of retrotransposition in Arabidopsis. Nature. 2009;461:427–30.

33. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature. 2001;411:212–4.

34. McClintock B. The significance of responses of the genome to challenge. Science. 1984;226:792–801.

35. Guerreiro MPG. What makes transposable elements move in the Drosophila genome? Heredity. 2012;108:461–8.

36. Enoki Y, Sakurai H. Diversity in DNA recognition by heat shock transcription factors (HSFS) from model organisms. FEBS Lett. 2011;585:1293–8.

37. Wilczek AM, Cooper MD, Korves TM, Schmitt J. Lagging adaptation to warming climate in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2014;111:7906–13.

38. Boyko A, Filkowski J, Kovalchuk I. Homologous recombination in plants is temperature and day-length dependent. Mut Res. 2005;572:73–83.

39. Pecinka A, Rosa M, Schikora A, Berlinger M, Hirt H, Luschnig C, et al. Transgenerational stress memory is not a general response in Arabidopsis. PLoS One. 2009;4:e5202.

40. Larionov A, Krause A, Miller W. A standard curve based method for relative real time pcr data processing. BMC Bioinformatics. 2005;6:62.

41. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg S. Tophat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.

42. Rawat V, Abdelsamad A, Pietzenuk B, Seymour DK, Koenig D, Weigel D, et al. Improving the annotation of *Arabidopsis lyrata* using RNA-seq data. PLoS One. 2015;10:e0137391.

43. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.

44. R Core Team. R: A language and environment for statistical computing. Vienna: Foundation for Statistical Computing; 2013. http://www.R-project.org/.

45. Quinlan AR, Hall IM. Bedtools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

46. Xu Z, Wang H. Ltr_finder: An efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–8.

47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

48. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 2012;40:D1178–86.

49. Edgar RC. Muscle: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

50. Bryant D, Moulton V. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 2004;21:255–65.

51. Huson DH. Splitstree: Analyzing and visualizing evolutionary data. Bioinformatics. 1998;14:68–73.

52. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23:254–67.

53. Thollesson M. Lddist: A perl module for calculating logdet pair-wise distances for protein and nucleotide sequences. Bioinformatics. 2004;20:416–8.

54. Kumar S, Stecher G, Tamura K. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.

55. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. Vista: Visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000;16:1046–7.

56. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. Vista: Computational tools for comparative genomics. Nucleic Acids Res. 2004;32:W273–9.

57. Hall TA. Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/nt. Nucleic Acids Symp Ser. 1999;41:95–8.

58. Bray N, Dubchak I, Pachter L. Avid: A global alignment program. Genome Res. 2003;13:97–102.

# Publikace 7

# Repair of DNA Damage Induced by the Cytidine Analog Zebularine Requires ATR and ATM in Arabidopsis<sup>OPEN</sup>

Chun-Hsin Liu,[a,1] Andreas Finke,[a,1] Mariana Díaz,[a] Wilfried Rozhon,[b] Brigitte Poppenberger,[b] Tuncay Baubec,[c] and Ales Pecinka[a,2]

[a] Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany
[b] Biotechnology of Horticultural Crops, Technische Universität München, 85354 Freising, Germany
[c] University of Zürich, 8057 Zürich, Switzerland

ORCID ID: 0000-0001-8474-6587 (T.B.)

**DNA damage repair is an essential cellular mechanism that maintains genome stability. Here, we show that the nonmethylable cytidine analog zebularine induces a DNA damage response in *Arabidopsis thaliana*, independent of changes in DNA methylation. In contrast to genotoxic agents that induce damage in a cell cycle stage-independent manner, zebularine induces damage specifically during strand synthesis in DNA replication. The signaling of this damage is mediated by additive activity of ATAXIA TELANGIECTASIA MUTATED AND RAD3-RELATED and ATAXIA TELANGIECTASIA MUTATED kinases, which cause postreplicative cell cycle arrest and increased endoreplication. The repair requires a functional STRUCTURAL MAINTENANCE OF CHROMOSOMES5 (SMC5)-SMC6 complex and is accomplished predominantly by synthesis-dependent strand-annealing homologous recombination. Here, we provide insight into the response mechanism for coping with the genotoxic effects of zebularine and identify several components of the zebularine-induced DNA damage repair pathway.**

## INTRODUCTION

Genome stability is frequently challenged by internal and external damaging factors, leading to formation of aberrant bonds, breakage, or cleavage of DNA (Britt, 1996). Genome damage is opposed by diverse surveillance mechanisms, with the DNA damage repair machinery playing the central role (Kolodner et al., 2002). Depending on the type of DNA damage, the plant induces different repair pathways, with evolutionarily conserved kinases activating specific repair processes. ATAXIA TELANGIECTASIA MUTATED (ATM) signals the existence of DNA double-strand breaks, and ATAXIA TELANGIECTASIA MUTATED AND RAD3-RELATED (ATR) signals the presence of single-stranded DNA, mostly at stalled replication forks (Cimprich and Cortez, 2008). This induces a cascade of responses affecting cell cycle progression (De Schutter et al., 2007) and activates the corresponding DNA damage repair effectors (Garcia et al., 2003; Culligan et al., 2006).

Recent studies have demonstrated the connections between DNA damage repair, genome integrity, and chromatin control (Downey and Durocher, 2006). Functional chromatin is important for genome stability, as loss of DNA methylation or defective nucleosome assembly increases sensitivity to genotoxic stress and alters homologous recombination (HR) frequencies in *Arabidopsis thaliana* (Kirik et al., 2006; Melamed-Bessudo and Levy, 2012; Rosa et al., 2013). However, higher frequency of somatic HR can be induced by zebularine, the nonmethylable cytidine analog used

for interference with transcriptional gene silencing (TGS) of various genetic elements (Zhou et al., 2002; Egger et al., 2004; Baubec et al., 2009, 2014; Pecinka et al., 2009). In addition, zebularine and 5-azacytidine (a less stable cytidine analog) treatments affect plant growth more severely than mutations in the genes responsible for DNA methylation, e.g., the SWI2/SNF2 class chromatin remodeling factor *DECREASED DNA METHYLATION1* (*DDM1*) (Baubec et al., 2009). This contrasts with the weaker DNA demethylation induced by zebularine treatment compared with that in the *ddm1* mutants (Baubec et al., 2009) and suggests that toxicity of nonmethylable cytidine analogs, and not DNA demethylation, could cause the reduction of plant growth in the presence of zebularine.

Zebularine and 5-azacytidine have been described as suppressors of tumor growth and are frequently used in cancer treatment, where zebularine is preferred, in some cases, over 5-azacytidine because of its lower toxicity (Dote et al., 2005; Yang et al., 2013). This is most likely due to the extensive metabolism of zebularine into zebularine-deoxyphosphate-cholines and diphosphoethanolamine, which may reduce the amount of biologically active drug (Ben-Kasus et al., 2005). Up to 5% of total cytosines can be replaced by 5-azacytidine, but the rate of zebularine incorporation into genomic DNA seems to be much lower (Jones and Taylor, 1980; Ben-Kasus et al., 2005). Both drugs are bound by DNA METHYLTRANSFERASEs (DNMTs) and form nucleoprotein adducts (NPAs), which effectively deplete the DNMT pool (Egger et al., 2004). In vitro studies using synthetic oligonucleotides containing 5-azacytidine or zebularine revealed higher stability of NPAs when compared with DNMT bound to 5-methyl-deoxycytosine (Champion et al., 2010; Kiianitsa and Maizels, 2013). The data generated using 5-azacytidine and 5-azadeoxicytidine suggest that NPAs represent a physical barrier for enzymes sliding along the DNA molecule and are repaired by HR coupled with replication restart and nucleotide excision repair (Kuo et al., 2007; Salem et al., 2009).

Nucleoside analogs are frequently used in basic and medical research. However, their mode of action and spectrum of effects is not well understood. Using Arabidopsis as a model system, we show that administration of zebularine triggers a specific type of DNA damage response, which dominates over DNA methylation changes. Reduced DNA damage response in the DNMT triple mutant suggests zebularine-DNMT NPAs as the possible causal aberrations. Zebularine treatment extends the G2 phase of the cell cycle and promotes endoreplication. Activation of DNA damage repair of zebularine-induced lesions is additively mediated by ATR and ATM kinases, and the damage is repaired by HR with only a minor contribution of nucleotide excision repair (NER). Absence of higher level of DNA strand breaks upon zebularine treatment differentiated its effects from those of 5-azacytidine inducing large amount of DNA single-strand breaks. The STRUCTURAL MAIN-TENANCE OF CHROMOSOMES5 (SMC5)-SMC6 complex plays an essential role in the repair of zebularine-induced DNA damage.

## RESULTS

### Transcriptional Activation of DNA Damage Repair Genes by Zebularine Treatment

To understand the effects of zebularine treatment, we used RNA-sequencing to perform genome-wide transcriptome analysis of dissected shoot apices of 12-d-old wild-type Arabidopsis plants treated with 20 μM zebularine for 24 h (short) and 5 d (long). Short and long zebularine treatment caused significant (adjusted P value < 0.05) upregulation of 31 and 678 genes and downregulation of 12 and 392 genes, respectively (Figure 1A, Table 1; Supplemental Data Set 1). The RNA-sequencing results for 12 significantly up- or downregulated genes were validated by reverse transcription-quantitative PCR (RT-qPCR) and revealed >75% agreement between both methods, including for key DNA damage repair genes (Supplemental Table 1). Only 38.7% of up- and 50% of down-regulated genes after short zebularine treatment overlapped with the set of genes differentially transcribed after long exposure (Figure 1A, Table 1). This indicated duration-dependent contrasting effects of zebularine treatment on the Arabidopsis transcriptome.

To identify how many of the zebularine up- or downregulated genes are targets of TGS, we compared our data to the RNA-sequencing data set of *ddm1* plants (Zemach et al., 2013). No overlap was found for short zebularine treatment and only four out of 908 genetic elements upregulated in *ddm1* were also signifi-cantly upregulated after the long zebularine treatment (TE gene AT1G42050; *MuDr* AT2G15810, *LINE1-6* AT3G28915, and *Gypsy-like* AT5G35057; Figure 1A). Therefore, <1% of the zebularine upregulated genes in shoot apices are TGS targets. A functional annotation analysis (TAIR10) of the 31 genes induced by the short zebularine treatment revealed that 32.3% are linked to DNA me-tabolism and DNA damage repair, e.g., the genes encoding the RIBONUCLEOTIDE REDUCTASE (RNR) complex subunits RNR1 and TSO2, and the genes *BREAST CANCER SUSCEPTIBLE1* (*BRCA1*), *RAS ASSOCIATED WITH DIABETES51* (*RAD51*), or *SIAMESE-RELATED7* (Table 1). Several additional DNA damage repair genes, including *GAMMA-IRRADIATION AND MITOMYCIN C INDUCED1* (*GMI1*), were significantly upregulated after the long zebularine treatment (Supplemental Data Set 1). To test whether

these mRNA changes represent a bona fide response to a DNA damage stimulus, we exposed plants to mitomycin C (MMC), a drug that induces DNA interstrand cross-links (Iyer and Szybalski, 1963; Tomasz, 1995). Short (24 h) 10 μM MMC treatment signifi-cantly up- and downregulated 815 and 579 genes, respectively, including numerous DNA damage repair genes (adjusted P value < 0.05; Figure 1B; Supplemental Data Set 2). Importantly, the sets of genes up- and downregulated in response to 24 h of zebularine exposure overlapped 93.1% (29 out of 31) and 91.7% (11 out of 12), respectively, with the MMC treatment (Figure 1B).

Prior to incorporation into DNA, zebularine undergoes modifi-cation in several steps (Ben-Kasus et al., 2005). This raises the question of the kinetics of the DNA damage response and its tissue specificity. To examine this, we used a *pGMI1::GUS* (β-glucuronidase) reporter line that allows the visualization of tis-sues with ongoing DNA damage repair (Böhmdorfer et al., 2011). The reporter lines were exposed to zebularine, MMC, and the ra-diomimetic drug bleocin. GUS was not detected in mock-treated plants, while 3 h of bleocin and 6 h of MMC or zebularine treatment were sufficient to obtain GUS staining in the shoot apices, petioles of the youngest leaves, and in the cotyledon vasculature (Figure 1C). Over time, the staining became more prominent in the entire true leaves and cotyledon vasculature. GUS was also detected in root apical meristems of MMC- and bleocin-treated, but not of zebularine-treated, samples. These results suggest a rapid in-duction of *GMI1* by zebularine and its different drug processing or stability in root and shoot apical meristem tissues. To assess the kinetics of transcriptional activation in more detail, we dissected shoot apices of mock- and drug-treated plants over the 24-h time series and validated *GMI1* activation by RT-qPCR (Figure 1D). However, the amount of transcript did not simply accumulate over time as observed in histochemical staining (Figure 1C), probably reflecting the higher stability of the GUS protein compared with *GMI1* mRNA. Other tested DNA damage repair genes, including those detected in our RNA-sequencing (*RAD51*, *BRCA1*, and *PARP2*) were also upregulated in response to zebularine with ki-netics and amplitudes similar to the MMC and bleocin treatments (Figure 1D). Hence, zebularine treatment leads to transcriptional upregulation of a specific set of DNA damage repair genes in shoot apical tissues, in a rapid and high amplitude manner.

### Zebularine-Triggered DNA Damage Response Is Independent of DNA Methylation Changes

Zebularine has been shown to reduce DNA methylation in a dose-dependent manner (Baubec et al., 2009). Therefore, the activation of DNA damage repair genes observed after 20 μM zebularine treatment may be caused by DNA demethylation. We identified methylated DNA regions <1 kb upstream of *TSO2* and *RAD51*, two DNA damage repair genes activated by zebularine treatment (Supplemental Figures 1 and 2). Analysis of these regions by bi-sulfite sequencing in dissected shoot apices of mock, short, and long zebularine-treated plants revealed <5% reduction of DNA methylation (Figure 2A; Supplemental Data Sources 1 to 4). Simi-larly, we observed normal levels of DNA methylation at the *LINE1-6* retrotransposon (AT3G28915/AT3TE45385) identified as a com-mon target of zebularine and activation in *ddm1* mutants. DNA methylation was also maintained in the repetitive region upstream
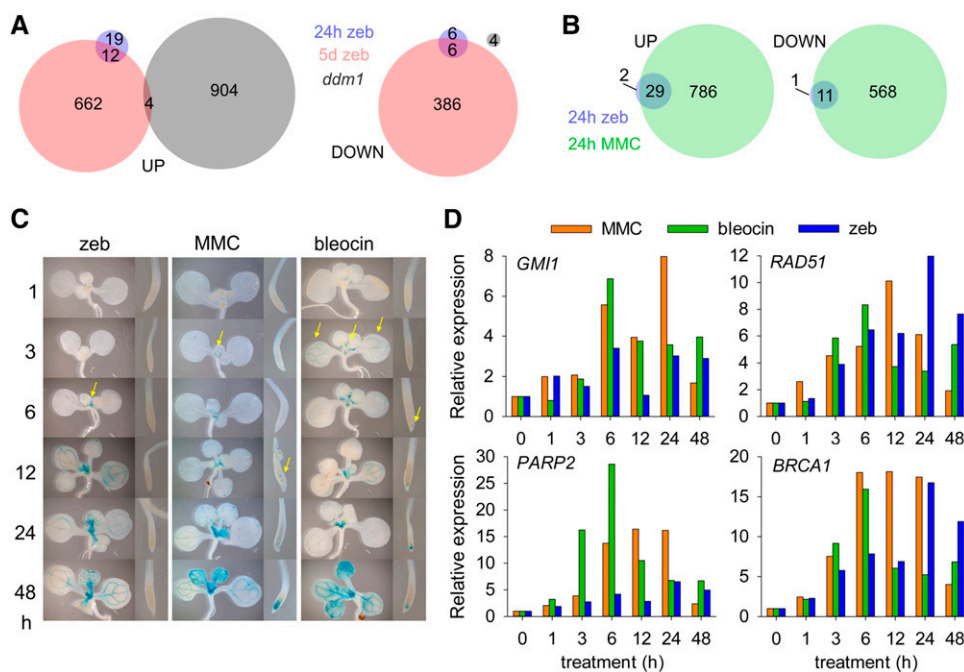
**Figure 1.** Zebularine Treatment Activates DNA Damage Repair Genes.

**(A)** Genes significantly up- or downregulated in response to 24 h (blue) and 5 d (pink) 20 μM zebularine (zeb) treatment of wild-type plants.

**(B)** Significantly up- and downregulated genes in response to 24 h zebularine (blue) and 24 h 10 μM MMC treatment (green).

**(C)** Histochemical staining of *pGMI1:GUS* reporter line after the specified hours of treatment with 20 μM zebularine, 10 μM MMC, and 100 nM bleocin. Representative rosettes and root tips are shown.

**(D)** RT-qPCR analysis of DNA damage repair marker genes *GMI1*, *RAD51*, *PARP2*, and *BRCA1* in dissected shoot apices after given hours of treatment with 20 μM zebularine, 10 μM MMC, and 100 nM bleocin. The bars represent a mean of mRNA levels from a pool of 5 to 10 seedlings in one biological replicate.

of the *SUPPRESSOR OF drm1 drm2 cmt3* (*SDC*) gene (Henderson and Jacobsen, 2008) upregulated by long zebularine treatment (Figures 2A and 2B; Supplemental Data Set 1). Hence, zebularine-induced upregulation of several genetic elements occurred without loss of DNA methylation. Recently, it has been shown that *SDC* can be activated by disturbed higher chromatin order structure in *MORC6* ATPase mutants (Moissiard et al., 2012). Because zebularine treatment leads to heterochromatin decondensation in Arabidopsis (Baubec et al., 2009), we tested whether disturbed chromatin structure in *morc6* is sufficient for induction of DNA damage repair genes. However, *SDC* but not *TSO2* and *RAD51* were activated in dissected apices of *morc6* plants (Supplemental Figure 3A). This suggests that disturbed heterochromatin structure alone is not sufficient to induce DNA damage repair response and that zebularine treatment interferes with at least two independent genome maintenance pathways. Furthermore, zebularine-induced transcriptional activation of DNA damage repair genes and TGS targets may occur without stable changes in DNA methylation.

Next, we tested the frequency of zebularine incorporation into plant genomic DNA. We grew Arabidopsis plants in medium containing 20 μM zebularine, which we refreshed every 3 d, for 14 d, and analyzed the amount of deoxyzebularine in genomic DNA using reverse-phase HPLC (RP-HPLC). Even with a detection limit at ∼1 deoxyzebularine per 5000 deoxycytosines (Supplemental

Figure 4), we could not detect deoxyzebularine incorporated into plant DNA. Although surprising, these data are generally in line with the low rate (∼0.1 pmol per μg DNA = ∼1 deoxyzebularine per ∼8000 deoxycytosines) of zebularine incorporation into DNA of mammalian cell lines (Ben-Kasus et al., 2005). This suggests that zebularine may not be efficiently and/or stably incorporated into DNA, in particular in Arabidopsis, a plant with very small meristems.

In vitro experiments with synthetic oligonucleotides revealed that DNMTs covalently bind to zebularine-containing DNA molecules (Champion et al., 2010). Since we could not detect zebularine directly in DNA, we tested whether the NPAs could cause DNA damaging effects by reducing the amount of available DNMTs. Due to strongly reduced fitness and pleiotropic effects of mutants in *DNA METHYLTRANSFERASE1* (Mathieu et al., 2007), we used *CHROMOMETHYLASE3* (*CMT3*), *DOMAINS REARRANGED METHYLTRANSFERASE1* (*DRM1*), and *DRM2* triple homozygous mutant (*ddc*) plants (Henderson and Jacobsen, 2008). We exposed *ddc* plants to mock treatment and 20 μM zebularine for 24 h and used RT-qPCR to measure mRNA levels of DNA damage repair genes. *TSO2*, *BRCA1*, *PARP2*, and *RAD51B* were 3.5- to 5.5-fold upregulated in response to zebularine in the wild type, whereas we observed <2-fold upregulation in zebularine-treated *ddc* plants (Figure 2B). Furthermore, zebularine-induced inhibition of root growth was significantly reduced in *ddc* compared with wild-type plants (*t* test, P < 0.05; Figure 2C; Supplemental Figure 3B).

**Table 1.** Genes Significantly Up- and Downregulated after Short (24 h) 20 μM Zebularine Treatment

| AGI Locus | Gene Annotation | Mock | | Zeb | | Log$_2$ Fold Change | Adjusted P Value | DDR |
|---|---|---|---|---|---|---|---|---|
| | | RPKM | ±SD | RPKM | ±SD | | | |
| Upregulated genes | | | | | | | | |
| At1g11580 | *METHYLESTERASE PCR A (PMEPCRA)* | 8.3 | 0.1 | 15.1 | 0.4 | 0.84 | 0.008 | |
| **At1g20750** | **RAD3-like** | **0.0** | **0.0** | **0.4** | **0.1** | **Infinite** | **0.004** | + |
| At1g48460 | Unknown protein | 12.5 | 0.5 | 20.2 | 1.2 | 0.68 | 0.049 | |
| At1g63660 | GMP synthase | 16.1 | 1.7 | 26.7 | 1.8 | 0.72 | 0.049 | + |
| At1g65310 | *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/ HYDROLASE17 (XTH17)* | 0.8 | 0.1 | 3.2 | 0.1 | 1.95 | 0.043 | |
| At1g70260 | *USUALLY MULTIPLE ACIDS MOVE IN AND OUT TRANSPORTERS36 (UMAMIT36)* | 2.7 | 0.4 | 6.5 | 0.4 | 1.23 | 0.009 | |
| At1g72440 | *SLOW WALKER2 (SWA2)* | 10.0 | 1.5 | 16.5 | 0.2 | 0.70 | 0.048 | |
| At1g75780 | *TUBULIN β-1 CHAIN (TUB1)* | 8.5 | 0.2 | 14.3 | 0.8 | 0.73 | 0.037 | |
| At1g78370 | *GLUTATHIONE S-TRANSFERASE TAU 20 (GSTU20)* | 368.0 | 12.6 | 721.3 | 74.7 | 0.95 | 0.000 | |
| **At2g21790** | **RIBONUCLEOTIDE REDUCTASE1 (RNR1)** | **23.7** | **2.3** | **40.5** | **1.4** | **0.76** | **0.001** | + |
| At2g40360 | *ARABIDOPSIS THALIANA PESCADILLO ORTHOLOG1 (ATPEP1)* | 18.9 | 2.7 | 31.9 | 0.4 | 0.74 | 0.021 | |
| At2g43100 | *ISOPROPYLMALATE ISOMERASE2 (IPMI2)* | 60.8 | 6.7 | 113.8 | 12.4 | 0.88 | 0.000 | |
| **At3g03780** | **METHIONINE SYNTHASE2 (MS2)** | **77.3** | **10.3** | **137.3** | **0.9** | **0.81** | **0.011** | |
| **At3g07800** | **THYMIDINE KINASE 1A (TK1A)** | **13.9** | **2.6** | **29.7** | **4.1** | **1.07** | **0.000** | + |
| **At3g13470** | **CHAPERONIN-60BETA2 (CPN60BETA2)** | **99.1** | **8.3** | **158.8** | **16.1** | **0.67** | **0.005** | |
| **At3g15950** | **NAI2** | **32.2** | **1.4** | **49.1** | **2.0** | **0.59** | **0.009** | |
| At3g16150 | *ASPARAGINASE B1 (ASPGB1)* | 2.9 | 0.1 | 8.3 | 0.3 | 1.48 | 0.007 | |
| **At3g19680** | **Protein of unknown function (DUF1005)** | **14.0** | **2.8** | **30.1** | **3.1** | **1.07** | **0.008** | |
| **At3g27060** | **TSO2** | **63.2** | **3.7** | **127.2** | **9.7** | **0.99** | **0.005** | + |
| **At3g27630** | **SIAMESE-RELATED7 (SMR7)** | **0.6** | **0.5** | **5.0** | **0.1** | **3.08** | **0.049** | + |
| **At3g54810** | **BLUE MICROPYLAR END3 (BME3)** | **19.6** | **0.8** | **30.5** | **1.0** | **0.62** | **0.024** | |
| At3g59670 | Unknown protein | 4.4 | 0.5 | 9.9 | 0.5 | 1.17 | 0.000 | + |
| **At4g21070** | **BREAST CANCER SUSCEPTIBILITY1 (BRCA1)** | **3.4** | **0.5** | **9.3** | **1.1** | **1.43** | **0.000** | + |
| **At4g22410** | **Ubiquitin C-terminal hydrolase protein** | **0.0** | **0.0** | **1.8** | **0.1** | **Infinite** | **0.048** | |
| **At4g22880** | **LEUCOANTHOCYANIDIN DIOXYGENASE (LDOX)** | **7.2** | **0.2** | **15.7** | **2.3** | **1.10** | **0.003** | |
| At4g31210 | DNA topoisomerase | 10.7 | 0.9 | 16.4 | 0.2 | 0.61 | 0.035 | + |
| At5g14200 | *ISOPROPYLMALATE DEHYDROGENASE1 (IMD1)* | 76.1 | 2.0 | 146.9 | 23.2 | 0.92 | 0.000 | |
| **At5g20850** | **RAS ASSOCIATED WITH DIABETES51 (RAD51)** | **3.3** | **0.3** | **7.3** | **0.8** | **1.11** | **0.043** | + |
| **At5g42800** | **DIHYDROFLAVONOL 4-REDUCTASE (DFR)** | **5.3** | **0.4** | **11.8** | **0.9** | **1.14** | **0.000** | |
| At5g52470 | *FIBRILLARIN1 (FIB1)* | 83.8 | 1.4 | 128.8 | 4.7 | 0.60 | 0.049 | |
| At5g55920 | *OLIGOCELLULA2 (OLI2)* | 12.0 | 2.8 | 22.9 | 0.1 | 0.92 | 0.011 | |
| Downregulated genes | | | | | | | | |
| **At1g28330** | **DORMANCY-ASSOCIATED PROTEIN1 (DYL1)** | **170.6** | **5.9** | **106.3** | **14.3** | **−0.71** | **0.022** | |
| **At1g35612** | **Transposable element gene** | **40.9** | **1.3** | **26.7** | **4.8** | **−0.64** | **0.037** | |
| At1g68050 | *FLAVIN-BINDING, KELCH REPEAT, F BOX1 (FKF1)* | 4.8 | 0.7 | 1.8 | 0.6 | −1.46 | 0.003 | |
| At2g21210 | *SMALL AUXIN UPREGULATED RNA6 (SAUR6)* | 63.1 | 0.3 | 31.8 | 5.7 | −1.02 | 0.049 | |
| **At2g33830** | **DORMANCY ASSOCIATED GENE2 (DRM2)** | **317.4** | **62.1** | **90.6** | **13.0** | **−1.83** | **0.000** | |
| **At2g42530** | **COLD REGULATED 15B (COR15B)** | **50.7** | **2.6** | **16.9** | **2.7** | **−1.59** | **0.005** | |
| **At3g05880** | **RARE-COLD-INDUCIBLE 2A (RCI2A)** | **144.0** | **4.0** | **89.7** | **9.6** | **−0.71** | **0.005** | |
| At3g62550 | Adenine nucleotide α-hydrolase-like | 80.3 | 2.3 | 48.1 | 5.6 | −0.75 | 0.003 | |
| At4g04330 | *HOMOLOG OF CYANOBACTERIAL RBCX1 (RBCX1)* | 55.2 | 4.9 | 33.8 | 5.1 | −0.72 | 0.049 | |
| At4g39090 | *RESPONSIVE TO DEHYDRATION19 (RD19)* | 241.8 | 1.9 | 163.8 | 7.4 | −0.58 | 0.008 | |
| At5g14780 | *FORMATE DEHYDROGENASE (FDH)* | 84.4 | 3.5 | 57.9 | 3.5 | −0.56 | 0.010 | |
| **At5g54190** | **PROTOCHLOROPHYLLIDE OXIDOREDUCTASE A (PORA)** | **10.4** | **0.2** | **5.3** | **0.2** | **−1.00** | **0.009** | |

Reads per kilobase per million reads (RPKM) are an average of two biological replicates ± SD. Adjusted P values were calculated using DESeq statistics in R. DNA damage repair (DDR) genes (TAIR10) are marked with a "+." Genes in bold were significantly up- or downregulated after a long (5 d) zebularine treatment.
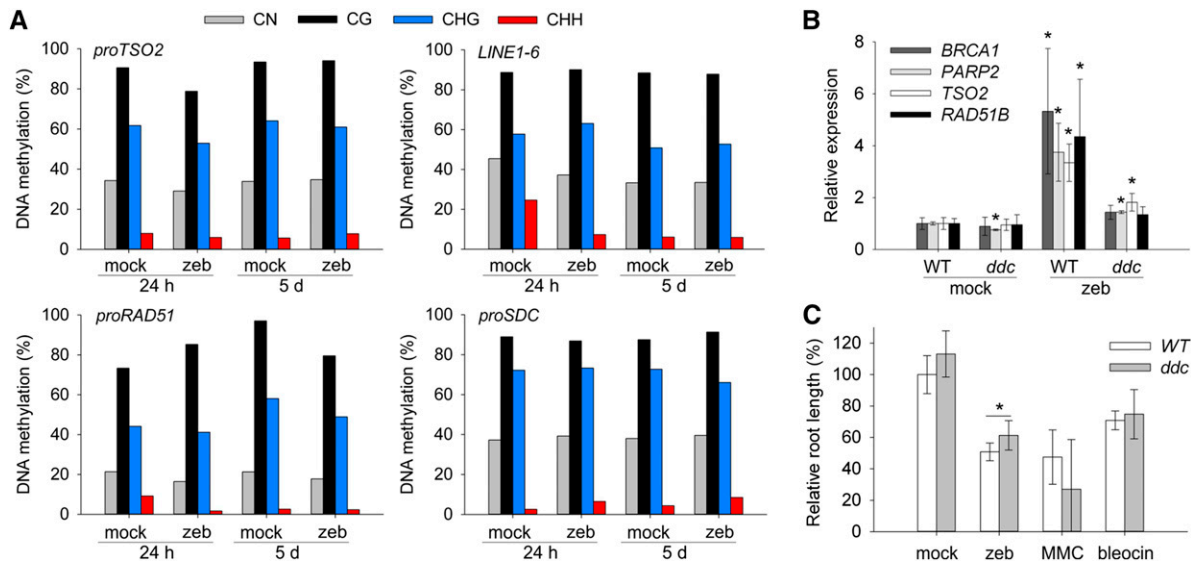
**Figure 2.** Zebularine Effects on DNA Methylation and Nucleoprotein Adduct Formation.

**(A)** Percentage of DNA methylation in dissected shoot apices based on bisulfite sequencing of 24 h and 5 d mock- and 20 μM zebularine (zeb)-treated samples. A minimum number of 12 reads per experimental point has been analyzed. Schematic view of the analyzed genomic regions is provided in Supplemental Figure 1.

**(B)** RT-qPCR measurement of DNA damage marker gene induction in the wild type (WT) and *drm1 drm2 cmt3* (*ddc*) triple mutant after 24 h treatment with mock and 20 μM zebularine normalized to *ACTIN7*. Error bars represent SD of three biological replicates and asterisks P < 0.05 in *t* test.

**(C)** Relative root length of wild-type and *ddc* plants in response to 20 μM zebularine, 15 μM MMC, or 50 nM bleocin treatment. Error bars represent SD of three biological replicates and asterisk P < 0.05 in *t* test.

Therefore, the DNMT-zebularine NPAs seem to be at least partly responsible for the DNA damage phenotypes and zebularine toxicity.

This indicates that zebularine incorporation into DNA is rare or unstable, the transcriptional activation of zebularine-induced targets occurs without stable DNA demethylation, and the DNA damage response is triggered at least partially by the zebularine-DNMT NPAs.

## ATR and ATM Redundantly Signal Repair of Zebularine-Induced DNA Damage

The >90% overlap between MMC and zebularine-induced mRNA changes suggests that the damage they induce is repaired by a pathway with at least some components in common. Interstrand DNA cross-linking activity of MMC causes stalled replication forks that are repaired by the ATR pathway (Culligan et al., 2004). Therefore, we performed RNA-sequencing of the shoot apices of *atr* mutant plants exposed to mock, 20 μM zebularine, and 10 μM MMC for 24 h and compared this with their effects on the wild type. In mock-treated *atr*, 227 and 119 genes were significantly up- and downregulated, respectively (Supplemental Figure 5A and Supplemental Data Set 3). This corresponded to 70 and 20 significantly enriched Gene Ontology term categories, respectively, pointing toward stress and immune responses (Supplemental Data Set 4). As *atr* plants were grown under conditions that did not induce stress in the wild type, this indicates that ATR prevents a hypersensitive reaction to the environment in Arabidopsis. A 24-h zebularine and MMC treatment of *atr* led to significant upregulation of 62 and 78 genes (29 common), respectively (Supplemental

Figure 5A). In total, 363 and 421 genes (225 overlapping) were significantly downregulated in *atr* in response to zebularine and MMC treatment, respectively (Supplemental Figure 5A). This confirms the role of ATR as a positive regulator of transcription in response to stress. Importantly, only four genes were commonly upregulated and two downregulated in zebularine-treated wild type and *atr*, suggesting that most of the transcriptional response to zebularine treatment is ATR dependent (Figure 3A). This was less pronounced for the MMC treatment, where 50% of upregulation (408 out of 815) and 61% of downregulation (353 out of 579) occurred in an ATR-independent manner (Figure 3B).

However, several genes upregulated in response to the zebularine treatment were also previously identified as ATM targets (Culligan et al., 2006). Therefore, we performed genetic studies to test for the involvement of both kinases in detoxifying zebularine-induced damage. Besides the reduced root length of *atr*, phenotypes of *atr* and *atm* were similar to the wild type on medium without zebularine. But both mutants had partially reduced growth on 20 μM zebularine (Figure 3C; Supplemental Figures 5B and 5C and Supplemental Tables 2 and 3). This resembled the phenotype obtained after bleocin treatment and contrasted with the MMC treatment, which caused an extreme hypersensitivity in *atr* and only weak sensitivity in *atm*. Next, we tested for potential functional redundancy of ATM and ATR in repair of zebularine-induced damage. Because the *atm atr* double mutants are sterile (Culligan et al., 2006), we phenotyped and genotyped a population of plants homozygous for *atr* (*ATR*[−/−]) and segregating for *atm* alleles (*ATM*[−/+]). In total 27.6% (16 out of 58) of plants were *atm atr* homozygous double mutants and corresponded to individuals with extreme hypersensitivity to the zebularine
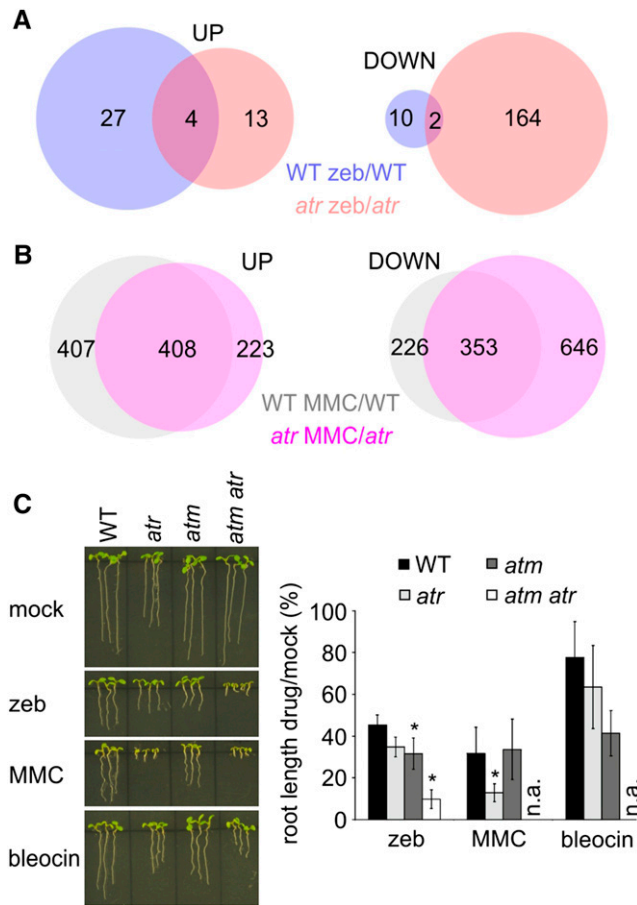
**Figure 3.** Both ATR and ATM Signal Repair of Zebularine-Induced Damage.

**(A)** Effects of zebularine-*atr* on gene transcript levels. Blue ovals in Venn diagrams show genes significantly up- or downregulated in response to short zebularine (zeb) exposure. Pink depicts genes significantly up- or downregulated in zebularine-treated relative to mock-treated *atr*. The genes in overlap are upregulated in response to zebularine independent of ATR. **(B)** MMC-*atr* effects on gene transcript amounts analyzed as described in **(B)**. **(C)** Representative phenotypes of wild-type, *atr*, *atm*, and *atm atr* double mutant root elongation on 20 μM zebularine, 15 μM MMC, and 50 nM bleocin. The graph shows quantitative root length data for individual genotypes. Asterisks indicate statistically significant (*t* test, P < 0.05), and error bars denote sd of three biological replicates. n.a., not analyzed.

treatment (Figure 3C; Supplemental Figures 5B and 5C). All *ATR*$^{-/-}$ *ATM*$^{-/+}$ plants were fully sensitive to MMC treatment due to *atr* single mutant hypersensitivity, and no fully sensitive *ATR*$^{-/-}$ *ATM*$^{-/+}$ plants were observed upon bleocin treatment (Figure 3C). These experiments provide molecular and genetic evidence for the additive role of ATR and ATM in signaling repair of zebularine-induced DNA damage.

## Zebularine-Induced DNA Damage Is Detoxified Predominantly by Intermolecular HR

Metazoan data suggest that activation of ATM may be triggered by both DNA strand breaks and disturbed chromatin structure (Bakkenist and Kastan, 2003). To test for the former, we performed

single cell electrophoresis (comet assays) using the alkaline/neutral protocol to detect both DNA single- and double-strand breaks. One-hour treatment of wild-type plants with 25 μg/mL bleocin resulted in 70% of DNA in comet tails, while only 10% of DNA was in the tails in the mock-treated sample (Figure 4A). However, the amount of DNA in the tail did not increase beyond mock levels during 24-h treatment with 800 μM zebularine (Figure 4A). This strongly suggests that even high zebularine concentrations over long time periods do not cause substantial fragmentation of the nuclear genome. This was further supported by the lack of zebularine hypersensitivity in mutants of nonhomologous end joining components *KU70* and *LIGASE* (*LIG4*), which were hypersensitive to bleocin treatment (Figures 4B and 4C; Supplemental Figure 6 and Supplemental Tables 2 and 3). We also tested effects of treatment with 5-azacytidine, another nonmethylable cytidine analog, on DNA integrity (Supplemental Figure 7). We observed significantly (*t* test, P < 0.01) more DNA in comet tails after 1 h of 100 and 200 μM 5-azacytidine treatment followed by alkaline/neutral comet assays. However, no increased tail DNA was found in neutral/neutral comet assays, indicative of DNA double-strand breaks. This suggests that 5-azacytidine treatment is associated with extensive DNA single-strand breakage, in contrast with zebularine treatment where no large amount of DNA strand breaks could be detected.

Strongly reduced growth of mutants in the genes encoding the CHROMATIN ASSEMBLY FACTOR1 components *FASCIATA1* (*FAS1*) and *FAS2* on zebularine suggested an additive effect of chemical and genetic interference with chromatin structure (Figure 4B; Supplemental Figures 6A to 6C and Supplemental Tables 2 and 3). Hence, ATM activation in response to zebularine treatment might occur via disturbed chromatin or DNA double helix structure.

ATR is activated by the presence of single-stranded DNA, typically at stalled replication forks (Cimprich and Cortez, 2008). Interference with the ATR pathway frequently leads to cell cycle prolongation or arrest (Culligan et al., 2004, 2006). We tested for zebularine-induced effects on the cell cycle using a cyclin-GUS (*pCYCB1;1::CYCB1;1:GUS*) reporter line (Colón-Carmona et al., 1999). This reporter protein is synthesized in G2 and degraded at the onset of mitosis. Under mock treatment conditions, the accumulation of cyclin-GUS can be observed in few root apical meristem cells (Figure 4D). Application of 10 μM MMC, 100 nM bleocin, or 20 μM zebularine led to time-dependent accumulation of GUS positive cells in root apical meristems. However, the strongest interference with the cell cycle occurred after MMC treatment followed by zebularine and bleocin treatments. Hence, zebularine-induced damage blocks progression of G2 to M phase. This block is weaker than MMC cross-links, but stronger than DNA double-strand breaks induced by bleocin, with the latter proposed to be repaired in a cell cycle stage-independent manner (Schubert et al., 2004)

To explore the detoxification mechanism of zebularine-induced DNA damage further, we analyzed the sensitivity of mutants of several DNA repair pathways. In bacteria, mutants defective in NER were hypersensitive to 5-azacytidine (Betham et al., 2010). Therefore, we exposed plants mutated in the *XERODERMA PIGMENTOSUM GROUP F* (*XPF*) gene, the endonuclease involved in NER and removal of nonhomologous overhangs in intramolecular homologous recombination events (Gaillard and Wood, 2001;
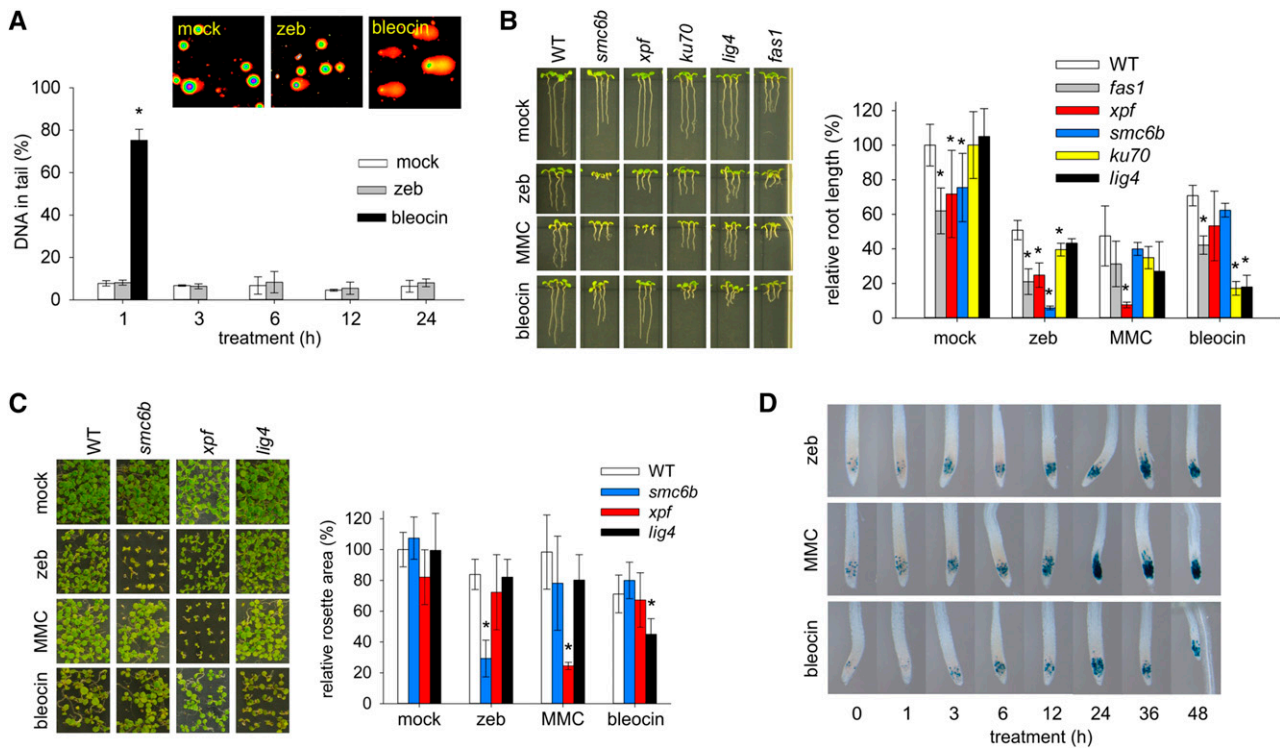
**Figure 4.** Zebularine Treatment Blocks Cell Cycle and Is Lethal for *smc6b* Plants.

**(A)** Analysis of DNA fragmentation in response to genotoxic treatment. Images of representative comet assays based on nuclei isolated from plants treated with mock, 800 μM zebularine (zeb), and 25 μg/mL bleocin for 1 h. The graph shows percentage of DNA in comet tail. Error bars indicate SD of means from three biological replicates, and asterisk marks statistically significantly different groups relative to mock control (*t* test; P < 0.05).
**(B)** and **(C)** Images show representative root length **(B)** and rosettes **(C)** of the wild type (WT) and mutants grown on mock, 20 μM zebularine, 15 μM MMC, and 100 nM bleocin for 7 and 15 d, respectively. Quantitative data presented in graphs are based on three to five biological replicates with the SD indicated by error bars. Statistically significant (P < 0.05) differences in *t* test are labeled by asterisk.
**(D)** Representative GUS-stained root tips of the cyclin-GUS reporter line after treatment with 20 μM zebularine, 10 μM MMC, and 100 nM bleocin for the given number of hours.

Dubest et al., 2002; Molinier et al., 2008; Yoshiyama et al., 2009), to zebularine and other drugs (Figures 4B and 4C; Supplemental Figures 6A and 6B and Supplemental Table 2). While the *xpf* plants were hypersensitive to MMC treatment, they showed much weaker sensitivity to zebularine. This suggests a minor role of NER and intramolecular homologous recombination in the repair of zebularine-induced DNA damage in Arabidopsis. Similar weak zebularine sensitivity was observed for *rad5a* plants (Supplemental Figure 8), indicating that repair of zebularine-induced damage does not occur via replication fork regression (Heyer et al., 2010). An opposite pattern was found for the mutants of *SMC6B*, which were hypersensitive to zebularine and only moderately sensitive to MMC treatment (Figures 4B and 4C; Supplemental Figures 6A, 6B, and 6D). SMC6B is the core component of the SMC5-SMC6 complex (Yan et al., 2013), which has been implicated in DNA damage repair processes in both animals and plants (Mengiste et al., 1999; Chiolo et al., 2011). In Arabidopsis, SMC6B (and presumably the entire SMC5-SMC6 complex) is required for the normal speed of lesion removal and frequency of HR (Mengiste et al., 1999; Hanin et al., 2000; Kozak et al., 2009; Watanabe et al., 2009).

We previously observed that zebularine strongly increases the frequency of somatic HR in Arabidopsis (Pecinka et al., 2009). However, a detailed analysis of this phenotype and comparison to other types of DNA damage was missing. We selected HR reporter lines 651 and IC9C with a similar basal recombination frequency, but differing as to the recombination mechanism (Puchta et al., 1995; Molinier et al., 2004). Line 651 contains a direct repeat of the recombination substrate and allows scoring of intramolecular HR by single strand annealing (SSA). In contrast, an inverted repeat reporter region in the IC9C line is repaired by intermolecular re-combination mechanism of synthesis-dependent strand annealing (SDSA). All drug treatments increased HR of both lines (Figure 5A; Supplemental Table 4). However, the damage induced by MMC and bleocin treatments was repaired predominantly by SSA, which was also the preferred HR pathway under non-stress conditions (Figure 5B). However, zebularine-induced damage was repaired significantly more frequently by SDSA than SSA when compared with other treatments (Fisher's exact test, P < 0.001), suggesting that intermolecular HR by SDSA is the favored HR mechanism to remove zebularine-induced damage. To test whether this SDSA occurs between sister chromatids or homologous chromosomes,

we analyzed plants homozygous and hemizygous for the IC9C reporter construct as described (Molinier et al., 2004). The homozygous and hemizygous IC9C plants contained on average 0.83 and 0.41 GUS spots per plant, respectively (Figure 5C; Supplemental Table 5). The number of GUS spots in hemizygous plants was ∼49.1% of that in homozygous ones, suggesting that virtually all zebularine-induced SDSA events occurred between sister chromatids.

To prevent mitosis with potentially aberrant chromosomes, some cells may undergo endoreplication (De Veylder et al., 2011). We used flow cytometry to measure the endoreplication frequency in cotyledons of drug-treated seedlings (Figure 5D; Supplemental Table 6). The cycle value (CV) of mock-treated plants was 1.36 and increased significantly to 1.52 in response to 10 μM zebularine treatment (relative CV = 111%; $t$ test, P < 0.05). Control treatments with 10 μM MMC and 50 nM bleocin yielded CVs of 1.51 (relative CV = 111%, P < 0.05) and 1.32 (relative CV = 97%), respectively. Hence, zebularine treatment increased the endoreplication level similarly to MMC, while DNA strand break induction did not. Next, we extended the analysis to zebularine and MMC hypersensitive mutants (Figure 5D; Supplemental Table 4). Mock-treated mutants were similar to the wild type, except for *atr* and *fas1*, which reached 93% (CV 1.26) and 113% (CV 1.54) of the wild type endoreplication level, respectively. The CV of *fas1* was further enhanced by zebularine and MMC treatments (CV 1.81 and 1.74, respectively, both P < 0.05 in $t$ test). For *atm* and *atm atr* plants, zebularine treatment increased endoreplication to 123.0 and 130.7% (CV 1.6 and 1.77; P = 0.386 and 0.024, respectively), while treatments with bleocin and MMC significantly increased endoreplication in both genotypes (Figure 5D). In contrast, response to either treatment did not increase significantly in *atr*, probably owing to large variation between biological replicates. The endoreplication levels of *smc6b* did not change significantly upon zebularine treatment (Figure 5D), despite its hypersensitivity. This contrasted with the effect of nonfunctional *XPF*, where hypersensitivity to MMC correlated with strongly increased cycle value (168%, CV 2.24, P < 0.05).

Collectively, this provides evidence that zebularine induces a complex type of lesion that affect the cell cycle, leading to significantly increased frequency of endoreplication. These lesions are repaired by HR with a crucial role of the SMC5-SMC6 complex.

## DISCUSSION

Chromatin mediates the proper regulation of transcription and maintains the stability of genetic information. Nonmethylable cytidine analogs are widely used in epigenetic and cancer research (Ben-Kasus et al., 2005; Yang et al., 2013; Baubec et al., 2014). However, their biological effects and the mechanism(s) of their action are not well understood (Pecinka and Liu, 2014). Here, we showed that exposure of Arabidopsis to zebularine induces a DNA damage response that is signaled additively by ATR and ATM and repaired through SDSA.

Approximately 32% of the genes upregulated by short zebularine treatment were associated with DNA damage repair and additional genes were induced after longer zebularine treatment. This contrasts with transcriptome analysis after 16 d of 5-azacytidine treatment in Arabidopsis, which revealed upregulation of a functionally
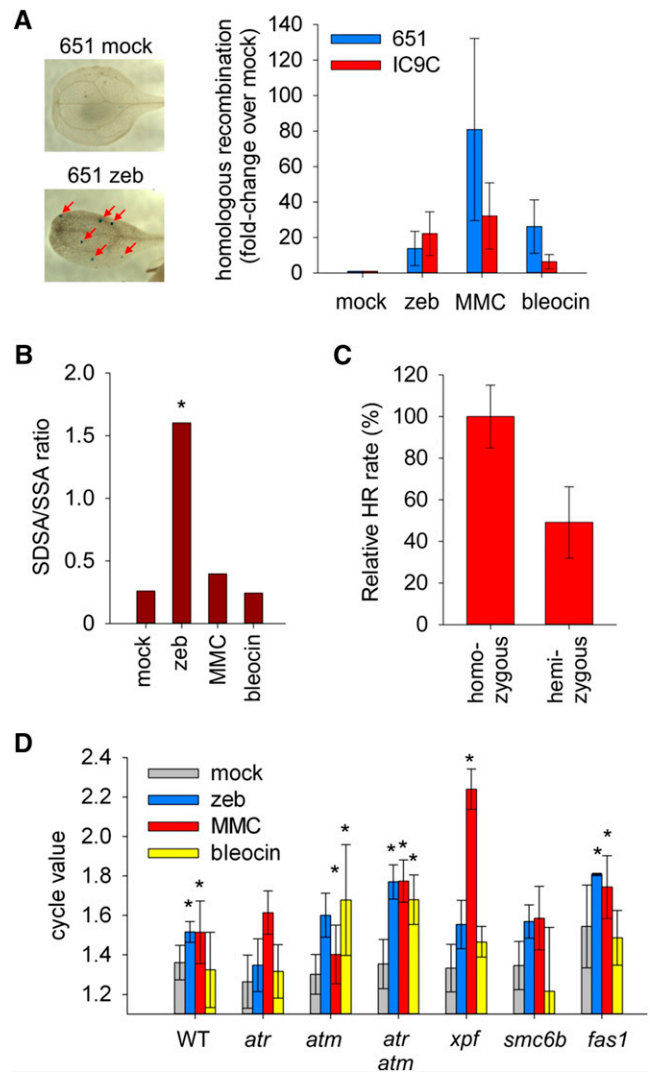


**Figure 5.** Zebularine Treatment Induces Endoreplication and Requires Repair by HR.

**(A)** HR assays. Left: representative cotyledons of mock and zebularine (zeb)-treated line 651. HR events, visible as blue dots, are indicated by red arrows. Right: HR frequency of SSA reporter line 651 and SDSA reporter line IC9C after 20 μM zebularine, 15 μM MMC, and 100 nM bleocin stress relative to mock treatment. Error bars denote sᴅ of three biological replicates.

**(B)** The ratio of SDSA versus SSA after different treatments. Asterisk indicates significant differences (P < 0.001) relative to mock treatment in Fisher's exact test.

**(C)** Average number of GUS spots in homozygous and hemizygous IC9C line after treatment with 20 μM zebularine. Error bars show sᴅ of four biological replicates.

**(D)** Mean cycle values of nuclei isolated from cotyledons of wild-type and mutant plants after 15 d of treatment with 10 μM zebularine, 10 μM MMC, and 50 nM bleocin. Error bars indicate sᴅ of three to five biological replicates, and asterisks denote statistically significant differences ($t$ test, P < 0.05).

diverse set of genes with no association to DNA damage repair (Chang and Pikaard, 2005). This is most likely due to differences in treatment length, stability, and biological effects of both drugs. In contrast, mRNA level changes induced by short zebularine treatment overlapped >90% with those induced by the alkylating agent MMC.

Because zebularine has been proposed to be an inhibitor of DNA methylation (Egger et al., 2004), we analyzed its genome-wide effects on the transcription of DNA methylation targets and also used bisulfite sequencing to analyze its effects on DNA methylation. Only four zebularine-activated genetic elements (<1%) were among the genes controlled by key DNA methylation factor DDM1 (Zemach et al., 2013). Another zebularine-activated TGS target included *SDC*, a gene under surveillance of DRM2 and CMT3 DNA methyltransferases and higher chromatin order established by MORC6 (Henderson and Jacobsen, 2008; Moissiard et al., 2012). However, *SDC* and the other three analyzed genes did not show DNA demethylation after the zebularine treatment. We cannot exclude DNA methylation changes in some specific genomic regions, but transcriptional activation of all analyzed genes occurred without loss of DNA methylation. This may be due to fast removal of zebularine, rapid DNA remethylation in apical meristems, or activation by reduced heterochromatin compaction (Baubec et al., 2009, 2014).

We were not able to detect deoxyzebularine in genomic DNA of treated plants with sensitivity of 1 deoxyzebularine per ~5000 deoxycytosines. Hence, the exact nature of zebularine-induced damage remains unknown. As a ribonucleotide, zebularine might be incorporated into RNA primers of Okazaki fragments and interfere with their removal. However, this model could not be experimentally tested owing to its technical difficulties. The reduced DNA damage response in *ddc* suggested that the damage is triggered at least partially by deoxy-zebularine-DNMT NPAs (Champion et al., 2010). NPAs (or DNA protein cross-links) are formed by the action of specific chemicals, UV radiation, or compromised activity of topoisomerases (Sheridan and Bishop, 2006; Stingele et al., 2014). Covalent binding of proteins to DNA is a common characteristic of NPAs that differentiates them from many other types of damage and requires specific repair components (Stingele et al., 2014). NPAs most likely represent a heterogeneous group due to different chemical or physical properties of their inducers and share some common features with other damaging agents. Our data also suggest fundamental differences in the nature of DNA damage induced by zebularine and 5-aza-cytidine, two structurally similar cytidine analogs.

Presumably, the nucleobase-like nature of zebularine allows its interference with genome stability only in a narrow window during DNA replication (Figure 5). As outlined above, this can be by incorporation into either newly synthesized DNA strands and/or RNA primers of Okazaki fragments. This contrasts with effect of MMC-induced interstrand cross-links, where damage is sensed before the replication fork; zebularine-induced damage most likely occurs later, during new strand synthesis. Hence, zebularine-induced DNA damage most likely occurs specifically after DNA strand separation. This activates the DNA damage repair machinery by additive functions of the kinases ATR and ATM. Previously, an additive role of ATM and ATR has been observed for the repair of DNA damage induced by ionizing radiation and in the course of meiosis (Culligan et al., 2006). However, our comparison to radiomimetic treatments

revealed that zebularine treatment interferes more strongly with DNA replication and does not cause extensive DNA strand break formation. Furthermore, zebularine treatment had a much stronger potential to increase endoreplication, which was similar to the replication-blocking agent. This creates a unique set of phenotypes that are not observed upon induction of DNA damage with other genotoxic agents and allows us to address the mechanism that repairs this damage.

To dissect repair pathways, we tested XPF, a component involved in NER and to some extent also in the SSA type of HR (Dubest et al., 2002; Molinier et al., 2008). The partial sensitivity of *xpf* shows that
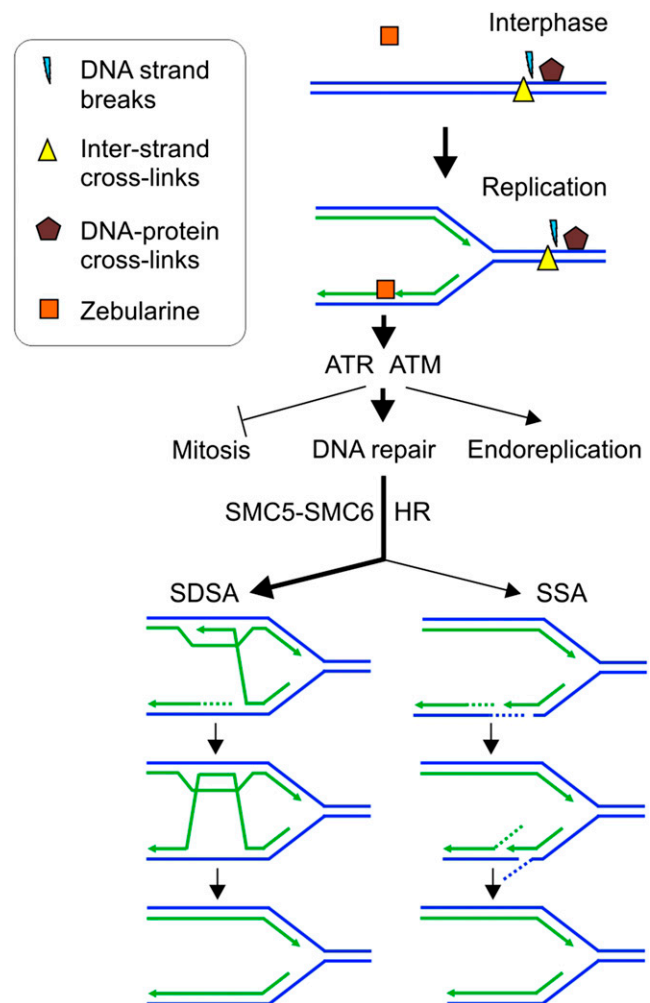


**Figure 6.** The Model of Zebularine-Induced Damage and Its Repair.

Most types of DNA damage, including DNA-protein cross-links, DNA strand breaks, or interstrand cross-links, can occur irrespectively of the cell cycle phase. In contrast, zebularine damage occurs during DNA replication in course of new DNA strand synthesis. This causes DNA damage stress, which suppresses cell division, promotes endoreplication, and activates DNA damage repair signaling by ATR and ATM activity. The repair depends strongly on SMC5-SMC6 activity and is pursued primarily by SDSA and to a smaller extent also SSA homologous recombination pathways.

a minor fraction of zebularine-induced damage is repaired by NER or SSA, which is also consistent with our HR data. This contrasts with the effects of 5-azacytidine, where NER is the dominant repair pathway in bacteria and humans (Salem et al., 2009; Orta et al., 2014). Because the *smc6b* mutant was more sensitive to zebularine than to other tested drugs, we suggest that the SMC5-SMC6 complex plays an essential role in the repair of zebularine-induced NPAs in Arabidopsis. We hypothesize that this could be either due to transcriptional deregulation of specific genes in *smc6b* or lack of DNA damage repair competence. It has been demonstrated that the SMC5-SMC6 complex functions as a facilitator of HR (Mengiste et al., 1999; Hanin et al., 2000; Watanabe et al., 2009) and its absence affects the speed of repair in Arabidopsis (Kozak et al., 2009). This is consistent with the proposed function of SMC5-SMC6 in controlling HR timing in DNA damage repair in *Drosophila melanogaster* (Chiolo et al., 2011) and also matches with the elevated frequency of somatic HR upon zebularine treatment (Pecinka et al., 2009).

The analysis of specific HR pathways revealed that SSA is a preferred HR pathway for repair of bleocin- and MMC-induced damage, while SDSA seems to be more important for repair of zebularine-induced damage. This is genetically supported by a minor role of XPF, an enzyme involved in HR by removing nonhomologous overhangs in SSA events (Dubest et al., 2002; Molinier et al., 2008). SSA can occur at both nonreplicated and replicated chromosomes, but SDSA only occurs at replicated chromosomes. By comparing plants allowing HR between sister chromatids and/or homologous chromosomes, we showed that zebularine-induced SDSA occurs strictly between sister chromatids. The lack of zebularine sensitivity of *rad5a* plants indicated the absence of replication fork regression (Heyer et al., 2010). Collectively, this suggests that zebularine-induced damage is removed after strand separation, during or shortly after the new strand synthesis (Figure 6). This further differentiates the zebularine effects from other DNA damaging agents and supports the presence of a specific repair strategy (Figure 6).

Zebularine is an anticancer agent that effectively suppresses growth of several types of tumors (Egger et al., 2004; Ben-Kasus et al., 2005; Yang et al., 2013). Mechanistically, this has been proposed to be due to its interference with DNA methylation and p53-dependent endoplasmic reticulum stress. Our data indicate an alternative mechanism based on the induction of specific DNA damage. Furthermore, Arabidopsis data suggest that this interference may be particularly effective for treatment of cells with deficient ATM and ATR functions.

## METHODS

### Plant Material

*Arabidopsis thaliana* wild type and mutants were in Columbia-0 background: *atm-1* (SALK_040423C), *atr-2* (SALK_032841C), *fas1* (Sail_662.D10), *fas2* (SALK_033228), *ku70* (SALK_123114C), *lig4* (SALK_044027C), *rad5a-2* (SALK_047150), *smc6b-1* (SALK_101968C), *smc6b-2* (SALK_135638), *smc6b-3* (Mengiste et al., 1999), and *xpf-3* (SALK_096156C). The *atr-2 atm-2* plants were identified in the *atr-2*$^{-/-}$ (SALK_032841C) and *atm-2*$^{+/-}$ (SALK_006953) segregating population. We also used cyclin-GUS containing the *pCYCB1;1::CYCB1;1:GUS* construct (Colón-Carmona et al., 1999) and *pGMI1:GUS* (Böhmdorfer et al., 2011). All mutants and reporter lines were used as homozygous lines unless stated otherwise. *smc6b-1* was used for experiments unless specified otherwise.

### Drug Treatments

The seeds were sterilized, evenly spread on sterile half-strength Murashige and Skoog (1/2 MS) medium with or without zebularine (Sigma-Aldrich), MMC (Duchefa Biochemie), and bleocin (Calbiochem) in concentrations specified in the text and grown at 16 h light:8 h dark at 21°C. For RNA-sequencing, RT-qPCR, and reporter analysis, plants were grown for 7 d on solid 1/2 MS medium and then transferred to control 1/2 MS plates or freshly prepared drug plates (Figures 1A, 1B, 2, 3, 4B, 4C, and 5; Supplemental Figures 3, 5, 6, and 8) or liquid media (Figures 1C, 1D, 4A, and 4D, Supplemental Figures 4 and 7) for specified times. For root elongation assays, 7-d-old plants grown continuously on mock and drug containing solid media were used. Fifteen-day-old plants grown under the same conditions were used for rosette area measurements and endoreplication analysis. RNA-sequencing was performed on dissected shoot apices of 12-d-old plants grown on solid media.

### Nucleic Acid Isolation, cDNA Synthesis, and RNA-Sequencing

DNA was extracted using the DNeasy kit (Qiagen) or Nucleon Phytopure kit (GE Healthcare). RNA was extracted using the RNeasy kit (Qiagen) with on-column DNase I (Roche) treatment. cDNA for quantitative PCR experiments was synthesized from 1 μg RNA per sample with Revert Aid H-Minus First Strand cDNA synthesis kit using the oligo-d(T) primer (Thermo Scientific). The purity of cDNA was monitored by PCR with an intron-spanning primer pair.

RNA sequencing was performed with two biological replicates per experimental point. The libraries were prepared from 1 μg total RNA with RNA integrity number >7.8 (Bioanalyzer; Agilent) using TruSeq RNA kit (Illumina) and sequenced as 100-bp single-end reads on HiSeq2500 (Illumina). Reads were trimmed and low-quality reads filtered with FAST-X tools (http://hannonlab.cshl.edu/fastx_toolkit/) using custom made scripts. This yielded an average of 15 million high-quality reads per library. The reads were mapped to the TAIR10 Arabidopsis reference genome using Tophat2 (Kim et al., 2013) with default settings. The coverage of individual genes was retrieved with the Qualimap from the set of uniquely mapped reads and significance (adjusted P value < 0.05) of mRNA level changes estimated with the DEseq package (Anders and Huber, 2010) in R. Venn diagrams were drawn using the venneuler package in R. Publicly available *ddm1* transcriptional data from the Gene Expression Omnibus data set GSE41302 (Zemach et al., 2013) were analyzed in the same way.

### Primers

Primers used in this study are provided in Supplemental Table 6.

### DNA Methylation Analysis

Approximately 120 ng of genomic DNA extracted from shoot apices of 15 seedlings was bisulfite treated using the EZ DNA methylation-Gold kit (Zymo Research). Desired fragments were PCR amplified from 1 μL of converted DNA and cloned into the pJET1.2 vector using the CloneJET PCR cloning kit (Thermo Scientific). At least 12 clones were analyzed per condition. Individual bisulfite sequencing reads used for analysis of DNA methylation are provided as Supplemental Data Sources 1 to 4.

### Comet Assays

Ten-day-old plants were transferred from 1/2 MS solid to liquid media containing no drug (mock), 25 μg/mL bleocin, 800 μM zebularine, and 100 or 200 μM 5-azacytidine for the specified times. Afterward, nuclei were isolated from entire seedlings and alkaline/neutral or neutral/neutral comet

assays were performed using the CometAssay kit (Trevigen) with the following modifications: The nuclei lysis was reduced to 5 min, unwinding to 10 min and electrophoresis to 6 min. Preparations were stained with Sybr Gold, and images were captured with a Zeiss Axio Imager A2 epifluorescence microscope equipped with Axiocam HRc camera. A total of 100 to 150 comets per experimental point were analyzed with CometScore (Tritek).

### GUS Staining and Endoploidy Analysis

GUS histochemical staining was performed as described (Baubec et al., 2009). Images were acquired using MZ16 FA stereomicroscope equipped with DFC490 CCD camera (both Leica). For endoploidy analysis, cotyledons were dissected, chopped with a razor blade in 300 μL extraction buffer (Partec), filtered through 30-μm nylon mesh, stained with 900 to 1800 μL CyStain dye (Partec), and analyzed with PAS I ploidy analyzer (Partec). The endopolyploidy cycle value was calculated using the formula: CV = ((n 2C*0) + (n 4C*1) + (n 8C+2) + (n 16C*3) + (n 32C*4)) / (n 2C + n 4C + n 8C + n 16C + n 32C), where n = number of counts per given C-value content.

### Quantitative PCR

The RT-qPCR was performed using 1 μL cDNA per 10-μL reaction with the SensiMix kit (PeqLab) on an CFX384 instrument (Bio-Rad). Fold changes were calculated relative to mock-treated controls using the standard curve method.

### Root Elongation and Rosette Area Measurements

For root length assay, plants were grown for 7 d on control and drug containing media, then carefully taken out using forceps without breaking roots and stretched on agar plates. Rosette area measurements were performed in independent experiments with 15-d-old plants. Plants were photographed with a D90 digital camera (Nikon). For rosette area measurements, color photographs were converted into binary mode using ImageJ (http://rsbweb.nih.gov/ij/). Both types of traits were then measured using ImageJ calibrated with an internal size control. Sensitivity to the DNA damaging agent in individual replicates was determined by calculating mean(treatment)/mean (mock). The roots and rosettes of at least 10 plants per genotype and treatment were measured per each of the three biological replicates.

### HR Assays

The 651 and IC9C reporter lines (Puchta et al., 1995; Molinier et al., 2004) were grown in liquid 1/2 MS media with or without drug treatment for 14 d, with the medium being replenished every 3 to 4 d. GUS staining was performed as described (Pecinka et al., 2009), and the number of GUS spots was examined under a stereomicroscope (Leica).

### *RP-HPLC*

DNA samples of zebularine- and mock-treated plants were prepared using the Plant DNA MaxiPrep kit (Qiagen). Two to six micrograms of DNA per sample was treated with DNase I and Nuclease P1 and subsequently with alkaline phosphatase to obtain the free dNs as described previously (Rozhon et al., 2008). The dNs composition was subsequently analyzed by RP-HPLC using a Nucleodur C18ec 100-5 125 × 4.6 mm column and a gradient starting with 98% eluent A (20 mM HCOOH set with NaOH to pH 4.0 in water) and 2% eluent B (20 mM HCOOH set with NaOH to pH 4.0 in 30% acetonitrile) at a flow rate of 0.8 mL/min. The concentration of eluent B was linearly increased to 5% within 7 min and subsequently to 50% within another 13 min. Finally, the initial settings were applied and the column equilibrated for 9.5 min prior injection of the next sample. Fluorescence of deoxyzebularine was detected at an excitation wavelength of 300 nm and an emission wavelength of 370 nm. UV absorbance was recorded at 277 nm.

### Supplemental Data

**Supplemental Figure 1.** DNA methylation analyzed regions.

**Supplemental Figure 2.** DNA sequences of genomic regions analyzed by bisulfite sequencing.

**Supplemental Figure 3.** Comparison of zebularine, *morc6*, and *ddc* phenotypes.

**Supplemental Figure 4.** Reverse-phase high performance liquid chromatography analysis of zebularine incorporation into genomic DNA.

**Supplemental Figure 5.** Rosette area of *atm*, *atr*, and *atm atr* under genotoxic stress.

**Supplemental Figure 6.** Mutant growth under genotoxic stress.

**Supplemental Figure 7.** 5-Azacytidine treatment causes DNA single-strand breaks.

**Supplemental Figure 8.** Phenotype of *rad5a* in zebularine root assay.

**Supplemental Table 1.** Validation of RNA-sequencing.

**Supplemental Table 2.** Relative root length (%) of mutants and the wild type treated by mock, zebularine, MMC, and bleocin.

**Supplemental Table 3.** Relative rosette area (%) of mutants and the wild type treated by mock, zebularine, MMC, and bleocin.

**Supplemental Table 4.** GUS spot numbers in HR reporter lines 651 and IC9C.

**Supplemental Table 5.** Number of GUS spots in IC9C homozygous and hemizygous plants after 20 μM zebularine treatment.

**Supplemental Table 6.** Cycle values after treatment with genotoxic stress.

**Supplemental Table 7.** PCR primers used in this study.

**Supplemental Data Set 1.** Genes significantly up- and downregulated after long (5 d) 20 μM zebularine treatment.

**Supplemental Data Set 2.** Genes significantly up- and downregulated after short (24 h) 10 μM mitomycin C treatment.

**Supplemental Data Set 3.** mRNA level changes in mock-, zebularine-, and mitomycin C-treated *atr* mutant.

**Supplemental Data Set 4.** Gene Ontology terms significantly enriched for the sets of genes up- and downregulated in mock-treated *atr* plants.

**Supplemental Data Source 1.** Bisulfite sequencing reads of *TSO2* promoter.

**Supplemental Data Source 2.** Bisulfite sequencing reads of *RAD51* promoter.

Supplemental Data Source 3. Bisulfite sequencing reads of *LINE1-6*.

Supplemental Data Source 4. Bisulfite sequencing reads of *SDC* promoter.

## REFERENCES

**Anders, S., and Huber, W.** (2010). Differential expression analysis for sequence count data. Genome Biol. **11:** R106.

**Bakkenist, C.J., and Kastan, M.B.** (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature **421:** 499–506.

**Baubec, T., Finke, A., Mittelsten Scheid, O., and Pecinka, A.** (2014). Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. EMBO Rep. **15:** 446–452.

**Baubec, T., Pecinka, A., Rozhon, W., and Mittelsten Scheid, O.** (2009). Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. Plant J. **57:** 542–554.

**Ben-Kasus, T., Ben-Zvi, Z., Marquez, V.E., Kelley, J.A., and Agbaria, R.** (2005). Metabolic activation of zebularine, a novel DNA methylation inhibitor, in human bladder carcinoma cells. Biochem. Pharmacol. **70:** 121–133.

**Betham, B., Shalhout, S., Marquez, V.E., and Bhagwat, A.S.** (2010). Use of Drosophila deoxynucleoside kinase to study mechanism of toxicity and mutagenicity of deoxycytidine analogs in *Escherichia coli.* DNA Repair (Amst.) **9:** 153–160.

**Böhmdorfer, G., Schleiffer, A., Brunmeir, R., Ferscha, S., Nizhynska, V., Kozák, J., Angelis, K.J., Kreil, D.P., and Schweizer, D.** (2011). GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in Arabidopsis. Plant J. **67:** 420–433.

**Britt, A.B.** (1996). DNA damage and repair in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47:** 75–100.

**Champion, C., Guianvarc'h, D., Sénamaud-Beaufort, C., Jurkowska, R.Z., Jeltsch, A., Ponger, L., Arimondo, P.B., and Guieysse-Peugeot, A.-L.** (2010). Mechanistic insights on the inhibition of c5 DNA methyltransferases by zebularine. PLoS ONE **5:** e12388.

**Chang, S., and Pikaard, C.S.** (2005). Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation. J. Biol. Chem. **280:** 796–804.

**Chiolo, I., Minoda, A., Colmenares, S.U., Polyzos, A., Costes, S.V., and Karpen, G.H.** (2011). Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. Cell **144:** 732–744.

**Cimprich, K.A., and Cortez, D.** (2008). ATR: an essential regulator of genome integrity. Nat. Rev. Mol. Cell Biol. **9:** 616–627.

**Colón-Carmona, A., You, R., Haimovitch-Gal, T., and Doerner, P.** (1999). Technical advance: spatio-temporal analysis of mitotic activity with a labile cyclin-GUS fusion protein. Plant J. **20:** 503–508.

**Culligan, K., Tissier, A., and Britt, A.** (2004). ATR regulates a G2-phase cell-cycle checkpoint in *Arabidopsis thaliana.* Plant Cell **16:** 1091–1104.

**Culligan, K.M., Robertson, C.E., Foreman, J., Doerner, P., and Britt, A.B.** (2006). ATR and ATM play both distinct and additive roles in response to ionizing radiation. Plant J. **48:** 947–961.

**De Schutter, K., Joubès, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., Van Der Schueren, E., Beeckman, T., Kushnir, S., Inzé, D., and De Veylder, L.** (2007). Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. Plant Cell **19:** 211–225.

**De Veylder, L., Larkin, J.C., and Schnittger, A.** (2011). Molecular control and function of endoreplication in development and physiology. Trends Plant Sci. **16:** 624–634.

**Dote, H., Cerna, D., Burgan, W.E., Carter, D.J., Cerra, M.A., Hollingshead, M.G., Camphausen, K., and Tofilon, P.J.** (2005). Enhancement of *in vitro* and *in vivo* tumor cell radiosensitivity by the DNA methylation inhibitor zebularine. Clin. Cancer Res. **11:** 4571–4579.

**Downey, M., and Durocher, D.** (2006). Chromatin and DNA repair: the benefits of relaxation. Nat. Cell Biol. **8:** 9–10.

**Dubest, S., Gallego, M.E., and White, C.I.** (2002). Role of the AtRad1p endonuclease in homologous recombination in plants. EMBO Rep. **3:** 1049–1054.

**Egger, G., Liang, G., Aparicio, A., and Jones, P.A.** (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature **429:** 457–463.

**Gaillard, P.-H.L., and Wood, R.D.** (2001). Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. Nucleic Acids Res. **29:** 872–879.

**Garcia, V., Bruchet, H., Camescasse, D., Granier, F., Bouchez, D., and Tissier, A.** (2003). AtATM is essential for meiosis and the somatic response to DNA damage in plants. Plant Cell **15:** 119–132.

**Hanin, M., Mengiste, T., Bogucki, A., and Paszkowski, J.** (2000). Elevated levels of intrachromosomal homologous recombination in Arabidopsis overexpressing the *MIM* gene. Plant J. **24:** 183–189.

**Henderson, I.R., and Jacobsen, S.E.** (2008). Tandem repeats upstream of the Arabidopsis endogene *SDC* recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev. **22:** 1597–1606.

**Heyer, W.-D., Ehmsen, K.T., and Liu, J.** (2010). Regulation of homologous recombination in eukaryotes. Annu. Rev. Genet. **44:** 113–139.

**Iyer, V.N., and Szybalski, W.** (1963). A molecular mechanism of mitomycin action: linking of complementary DNA strands. Proc. Natl. Acad. Sci. USA **50:** 355–362.

**Jones, P.A., and Taylor, S.M.** (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell **20:** 85–93.

**Kiianitsa, K., and Maizels, N.** (2013). A rapid and sensitive assay for DNA-protein covalent complexes in living cells. Nucleic Acids Res. **41:** e104.

**Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L.** (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14:** R36.

**Kirik, A., Pecinka, A., Wendeler, E., and Reiss, B.** (2006). The chromatin assembly factor subunit FASCIATA1 is involved in homologous recombination in plants. Plant Cell **18:** 2431–2442.

**Kolodner, R.D., Putnam, C.D., and Myung, K.** (2002). Maintenance of genome stability in *Saccharomyces cerevisiae.* Science **297:** 552–557.

**Kozak, J., West, C.E., White, C., da Costa-Nunes, J.A., and Angelis, K.J.** (2009). Rapid repair of DNA double strand breaks in *Arabidopsis thaliana* is dependent on proteins involved in chromosome structure maintenance. DNA Repair (Amst.) **8:** 413–419.

**Kuo, H.K., Griffith, J.D., and Kreuzer, K.N.** (2007). 5-Azacytidine induced methyltransferase-DNA adducts block DNA replication *in vivo*. Cancer Res. **67:** 8248–8254.

**Mathieu, O., Reinders, J., Čaikovski, M., Smathajitt, C., and Paszkowski, J.** (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell **130:** 851–862.

**Melamed-Bessudo, C., and Levy, A.A.** (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. Proc. Natl. Acad. Sci. USA **109:** E981–E988.

**Mengiste, T., Revenkova, E., Bechtold, N., and Paszkowski, J.** (1999). An SMC-like protein is required for efficient homologous recombination in Arabidopsis. EMBO J. **18:** 4505–4512.

**Moissiard, G., et al.** (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science **336:** 1448–1451.

**Molinier, J., Lechner, E., Dumbliauskas, E., and Genschik, P.** (2008). Regulation and role of Arabidopsis CUL4-DDB1A-DDB2 in maintaining genome integrity upon UV stress. PLoS Genet. **4:** e1000093.

**Molinier, J., Ries, G., Bonhoeffer, S., and Hohn, B.** (2004). Interchromatid and interhomolog recombination in *Arabidopsis thaliana.* Plant Cell **16:** 342–352.

**Orta, M.L., Höglund, A., Calderón-Montaño, J.M., Domínguez, I., Burgos-Morón, E., Visnes, T., Pastor, N., Ström, C., López-lázaro, M., and Helleday, T.** (2014). The PARP inhibitor Olaparib disrupts base excision repair of 5-aza-2′-deoxycytidine lesions. Nucleic Acids Res. **42:** 9108–9120.

**Pecinka, A., and Liu, C.-H.** (2014). Drugs for plant chromosome and chromatin research. Cytogenet. Genome Res. **143:** 51–59.

**Pecinka, A., Rosa, M., Schikora, A., Berlinger, M., Hirt, H., Luschnig, C., and Mittelsten Scheid, O.** (2009). Transgenerational stress memory is not a general response in Arabidopsis. PLoS ONE **4:** e5202.

**Puchta, H., Swoboda, P., Gal, S., Blot, M., and Hohn, B.** (1995). Somatic intrachromosomal homologous recombination events in populations of plant siblings. Plant Mol. Biol. **28:** 281–292.

**Rosa, M., Von Harder, M., Cigliano, R.A., Schlögelhofer, P., and Mittelsten Scheid, O.** (2013). The Arabidopsis SWR1 chromatin-remodeling complex is important for DNA repair, somatic recombination, and meiosis. Plant Cell **25:** 1990–2001.

**Rozhon, W., Baubec, T., Mayerhofer, J., Mittelsten Scheid, O., and Jonak, C.** (2008). Rapid quantification of global DNA methylation by isocratic cation exchange high-performance liquid chromatography. Anal. Biochem. **375:** 354–360.

**Salem, A.M.H., Nakano, T., Takuwa, M., Matoba, N., Tsuboi, T., Terato, H., Yamamoto, K., Yamada, M., Nohmi, T., and Ide, H.** (2009). Genetic analysis of repair and damage tolerance mechanisms for DNA-protein cross-links in *Escherichia coli.* J. Bacteriol. **191:** 5657–5668.

**Schubert, I., Pecinka, A., Meister, A., Schubert, V., Klatte, M., and Jovtchev, G.** (2004). DNA damage processing and aberration formation in plants. Cytogenet. Genome Res. **104:** 104–108.

**Sheridan, S., and Bishop, D.K.** (2006). Red-Hed regulation: recombinase Rad51, though capable of playing the leading role, may be relegated to supporting Dmc1 in budding yeast meiosis. Genes Dev. **20:** 1685–1691.

**Stingele, J., Schwarz, M.S., Bloemeke, N., Wolf, P.G., and Jentsch, S.** (2014). A DNA-dependent protease involved in DNA-protein crosslink repair. Cell **158:** 327–338.

**Tomasz, M.** (1995). Mitomycin C: small, fast and deadly (but very selective). Chem. Biol. **2:** 575–579.

**Watanabe, K., Pacher, M., Dukowic, S., Schubert, V., Puchta, H., and Schubert, I.** (2009). The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 complex promotes sister chromatid alignment and homologous recombination after DNA damage in *Arabidopsis thaliana.* Plant Cell **21:** 2688–2699.

**Yan, S., Wang, W., Marqués, J., Mohan, R., Saleh, A., Durrant, W.E., Song, J., and Dong, X.** (2013). Salicylic acid activates DNA damage responses to potentiate plant immunity. Mol. Cell **52:** 602–610.

**Yang, P.-M., Lin, Y.-T., Shun, C.-T., Lin, S.-H., Wei, T.-T., Chuang, S.-H., Wu, M.-S., and Chen, C.-C.** (2013). Zebularine inhibits tumorigenesis and stemness of colorectal cancer via p53-dependent endoplasmic reticulum stress. Sci. Rep. **3:** 3219.

**Yoshiyama, K., Conklin, P.A., Huefner, N.D., and Britt, A.B.** (2009). Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. Proc. Natl. Acad. Sci. USA **106:** 12843–12848.

**Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D.** (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell **153:** 193–205.

**Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J., and Hornby, D.P.** (2002). Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. J. Mol. Biol. **321:** 591–599.

**Repair of DNA Damage Induced by the Cytidine Analog Zebularine Requires ATR and ATM in Arabidopsis**

Chun-Hsin Liu, Andreas Finke, Mariana Díaz, Wilfried Rozhon, Brigitte Poppenberger, Tuncay Baubec and Ales Pecinka

This information is current as of July 12, 2015

| | |
|---|---|
| **References** | This article cites 57 articles, 24 of which can be accessed free at: http://www.plantcell.org/content/27/6/1788.full.html#ref-list-1 |
| **Permissions** | https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X |
| **eTOCs** | Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain |
| **CiteTrack Alerts** | Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain |
| **Subscription Information** | Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm |

# Publikace 8

# Pollen-Specific Activation of *Arabidopsis* Retrogenes Is Associated with Global Transcriptional Reprogramming[W][OPEN]

**Ahmed Abdelsamad[1] and Ales Pecinka[1,2]**

Max Planck Institute for Plant Breeding Research, Cologne DE-50829, Germany

**Duplications allow for gene functional diversification and accelerate genome evolution. Occasionally, the transposon amplification machinery reverse transcribes the mRNA of a gene, integrates it into the genome, and forms an RNA-duplicated copy: the retrogene. Although retrogenes have been found in plants, their biology and evolution are poorly understood. Here, we identified 251 (216 novel) retrogenes in *Arabidopsis thaliana*, corresponding to 1% of protein-coding genes. *Arabidopsis* retrogenes are derived from ubiquitously transcribed parents and reside in gene-rich chromosomal regions. Approximately 25% of retrogenes are cotranscribed with their parents and 3% with head-to-head oriented neighbors. This suggests transcription by novel promoters for 72% of *Arabidopsis* retrogenes. Many retrogenes reach their transcription maximum in pollen, the tissue analogous to animal spermatocytes, where upregulation of retrogenes has been found previously. This implies an evolutionarily conserved mechanism leading to this transcription pattern of RNA-duplicated genes. During transcriptional repression, retrogenes are depleted of permissive chromatin marks without an obvious enrichment for repressive modifications. However, this pattern is common to many other pollen-transcribed genes independent of their evolutionary origin. Hence, retroposition plays a role in plant genome evolution, and the developmental transcription pattern of retrogenes suggests an analogous regulation of RNA-duplicated genes in plants and animals.**

## INTRODUCTION

Gene duplications are an important factor in genome evolution, allowing for the functional diversification of genes (Flagel and Wendel, 2009; Innan and Kondrashov, 2010). Duplicated genes are generated by several DNA- and RNA-based mechanisms (Innan and Kondrashov, 2010; Sakai et al., 2011). Whole-genome DNA-based duplication (WGD) by polyploidization has occurred in the evolutionary history of all land plants and many animals (Dehal and Boore, 2005; De Smet et al., 2013). Since WGD amplifies the entire genome, it seems to be a solution toward major evolutionary and/or ecological challenges (Comai, 2005; Fawcett et al., 2009). However, WGDs do not alter protein stoichiometry in most cases; therefore, they may be relatively ineffective in situations where an increased amount of a single or a few specific proteins is required. In such situations, local DNA and RNA duplication mechanisms may be a more sophisticated solution. Local DNA duplications amplify individual genes or short chromosomal regions, presumably by an unequal crossing over mechanism (Zhang, 2003). In RNA-based duplication (retroposition), the mature mRNA of a protein-coding gene is reverse transcribed and integrated at an ectopic position

in the genome using retroviral or retrotransposon machinery (Kaessmann et al., 2009). Therefore, retroposition has a high potential to generate evolutionary innovations (e.g., by expressing genes in a new developmental context, generating chimeric genes with new functional domain combinations, or interspecific horizontal gene transfer) (Wang et al., 2006; Yoshida et al., 2010; Sakai et al., 2011). Relatively few studies have searched for retrogenes at the genome-wide scale in plants (Zhang et al., 2005; Wang et al., 2006; Zhu et al., 2009; Sakai et al., 2011). These studies have identified at most 0.38% of protein-coding genes as retrogenes, except for a study in maize (*Zea mays*) where low-stringency selection criteria were applied (Wang et al., 2006). In human (*Homo sapiens*), although 19.1% of all genes were identified as retrocopies, 82% of those contain premature stop codons. Therefore, 3.4% of all human genes are retrocopies producing putatively functional proteins (Marques et al., 2005; Pennisi, 2012). In rice (*Oryza sativa* subsp *japonica*), transcription was observed for two-thirds of retrogenes, indirectly suggesting that there may be a higher proportion of functional retrogenes in plants (Sakai et al., 2011).

Since retroposition duplicates only transcribed regions, it is expected to cause the loss of promoter sequences. This may represent a major bottleneck to retrogene evolutionary success. However, there are multiple possible mechanisms of retrogene promoter acquisition that have been demonstrated in individual examples (Kaessmann et al., 2009). Nevertheless, it is often not clear how frequent they are at the genome-wide scale. Recent studies in human and rice suggested that retroposition includes parental promoters (Okamura and Nakai, 2008).

Chromatin is an indispensable component that provides regulatory and protective functions to genetic information (reviewed in Li et al., 2007). Transcribed protein-coding genes are associated

---

with permissive chromatin marks. In contrast, transcriptionally repressed genes and repetitive elements are typically labeled by histone H3 Lysine 27 trimethylation (H3K27me3), histone H3 Lysine 9 dimethylation (H3K9me2), and/or high-density DNA methylation in all cytosine sequence contexts in plants (Roudier et al., 2011; Stroud et al., 2013). While H3K27me3 ensures tissue-specific developmental transcription (Lafos et al., 2011), the role of H3K9me2 and promoter DNA methylation is to minimize the activities of repetitive elements, which frequently include retrotransposons (Mosher et al., 2009; Slotkin et al., 2009; Ibarra et al., 2012). Retrogenes are generated by retro-transposon reverse transcriptases and represent duplicated copies. Therefore, they may become targets of epigenetic silencing by repressive chromatin. The association of retro-genes with specific chromatin states has been proposed (Boutanaev et al., 2002; Marques et al., 2005), but only a few have been characterized as to their chromatin states so far (Monk et al., 2011; Pei et al., 2012).

In flies and mammals, many retrogenes show testis-specific transcription (Marques et al., 2005; Vinckenbosch et al., 2006; Bai et al., 2008). This pattern is intriguing, and several explanatory models have been proposed (reviewed in Kaessmann et al., 2009; Kaessmann, 2010). First, it could originate from various chromatin modifications affecting chromosomes and leading to hypertranscription in meiotic and postmeiotic sper-matogenic cells. As a consequence of this global chromatin reorganization–induced transcription, some of the testis-transcribed retrogenes could also evolve testis-specific gene functions. The second, not mutually exclusive, hypothesis postulates that re-trogenes amplify in the germline tissues and insert preferentially into actively transcribed (open) chromatin. This creates a self-reinforcing loop where the retrogenes insert nearby or into germline-transcribed genes and consequently also would be germline-transcribed. The latter hypothesis is partially supported by observations in *Drosophila melanogaster* (Bai et al., 2008), but the tissue specificity in the transcription of plant retrogenes has not been clarified.

Here, we developed a search method that we used to identify 251 *Arabidopsis thaliana* retrogenes, 216 of which are novel. We use this set together with the retrogenes found previously to analyze retrogene and parent-specific features. We show that parents are usually ubiquitously transcribed, while retrogenes are mainly transcribed at low levels and in a stage-specific manner. Most *Arabidopsis* retrogenes acquired novel *cis*-regulatory elements at their integration sites, and introns significantly extend retrogene mRNA half-life. Importantly, throughout plant development, retrogenes show peaks of transcription in pollen. This pattern can also be observed for many lowly transcribed genes genome-wide and resembles retrogene transcription in the testis of animals.

## RESULTS

### *Arabidopsis* Retrogenes Are Capable of Repeated Retroposition and Occur in Gene-Rich Genomic Regions

We developed a bioinformatic method to identify retrogenes (Figure 1A). This was based on a genome-wide search for gene paralogy and retrogene-specific characteristics such as dif-ferential intron numbers relative to the parental gene and/or the presence of a poly(A) tail. The method was used to screen the genome of *Arabidopsis*, and in total, 251 retroposition events satisfying stringent quality criteria were identified (Supplemental Data Set 1). Among the retrogenes identified in our list, 36 were shared with two previous *Arabidopsis* genome-wide retrogene screens and 216 were novel (Figure 1B; Supplemental Data Sets 1 and 2) (Zhang et al., 2005; Zhu et al., 2009). The total number of retrogenes identified in all three studies is 309 (291 were considered for downstream analyses; Supplemental Data Sets 1 and 2), which corresponds to ~1% of *Arabidopsis* protein-coding genes and pseudogenes ($n$ = 27,416 and $n$ = 924, respectively).

Because our method combines multiple retrogene searches within intronless and intronized genes, it allows searching for potential secondary retropositions of retrogene transcripts. This revealed 12 retrogenes that served as templates for another round of retroposition (Supplemental Figure 1 and Supplemental Data Set 3). In these cases, the primary parent gave rise to the primary retrogene, whose mRNA served as the precursor for the secondary retrogene. The model where the primary parent gives rise directly to the secondary retrogene was not supported by the order of protein homologies and suggests retroposition of the retrogene transcript. Hence, 4.3% of *Arabidopsis* retrogenes underwent repeated retroposition without losing their protein-coding potential. In addition, we identified multiple-retrogene parents. In total, 22 parents gave rise to 54 retrocopies ($17 \times 2$, $3 \times 3$, $1 \times 4$, $1 \times 7$) and a maximum of seven retrocopies de-rived from a single parent (Supplemental Data Set 1). The ob-served frequency of multiple retropositions from the same gene is significantly higher than expected at random (Mann-Whitney-Wilcoxon [MWW] test, P < $2.2 \times 10^{-16}$), strongly arguing that the selection of parental mRNA is not random at least in some cases.

To explore whether retroposition occurs at specific genomic regions, we plotted the densities of all protein-coding genes, transposable elements (TEs), parents, and retrogenes over the five *Arabidopsis* chromosomes (Figure 1C). In agreement with published data (Arabidopsis Genome Initiative, 2000), TEs were enriched in pericentromeric regions and depleted from chro-mosome arms, while protein-coding genes showed the opposite pattern. Both retrogenes and parents had profiles similar to that of protein-coding genes, showing that they occur preferentially in gene-rich genomic regions (Figure 1C). To test for the asso-ciation of retrogenes and/or parents with TEs at the local scale, we estimated the frequency of genes with TEs in 1-kb intervals upstream and downstream of gene transcription start sites (TSSs) and transcription termination sites (TTSs). On average, there are fewer TEs upstream than downstream of genes. The frequency of TEs in TSS upstream regions of the genome-wide genes and retrogenes (17 and 22%, respectively) was not sig-nificantly different (Figure 1D). By contrast, parental genes with TEs in the first 2 kb upstream of the TSS were scarce relative to the whole genome ($\chi^2$ test, P < 0.05). Similarly, 25% of all genes and retrogenes contained TEs in the first 2 kb of the TTS downstream region, while it was only 17% for parents ($\chi^2$ test, P < 0.05 in the first 1 kb). This shows that retrogenes are not
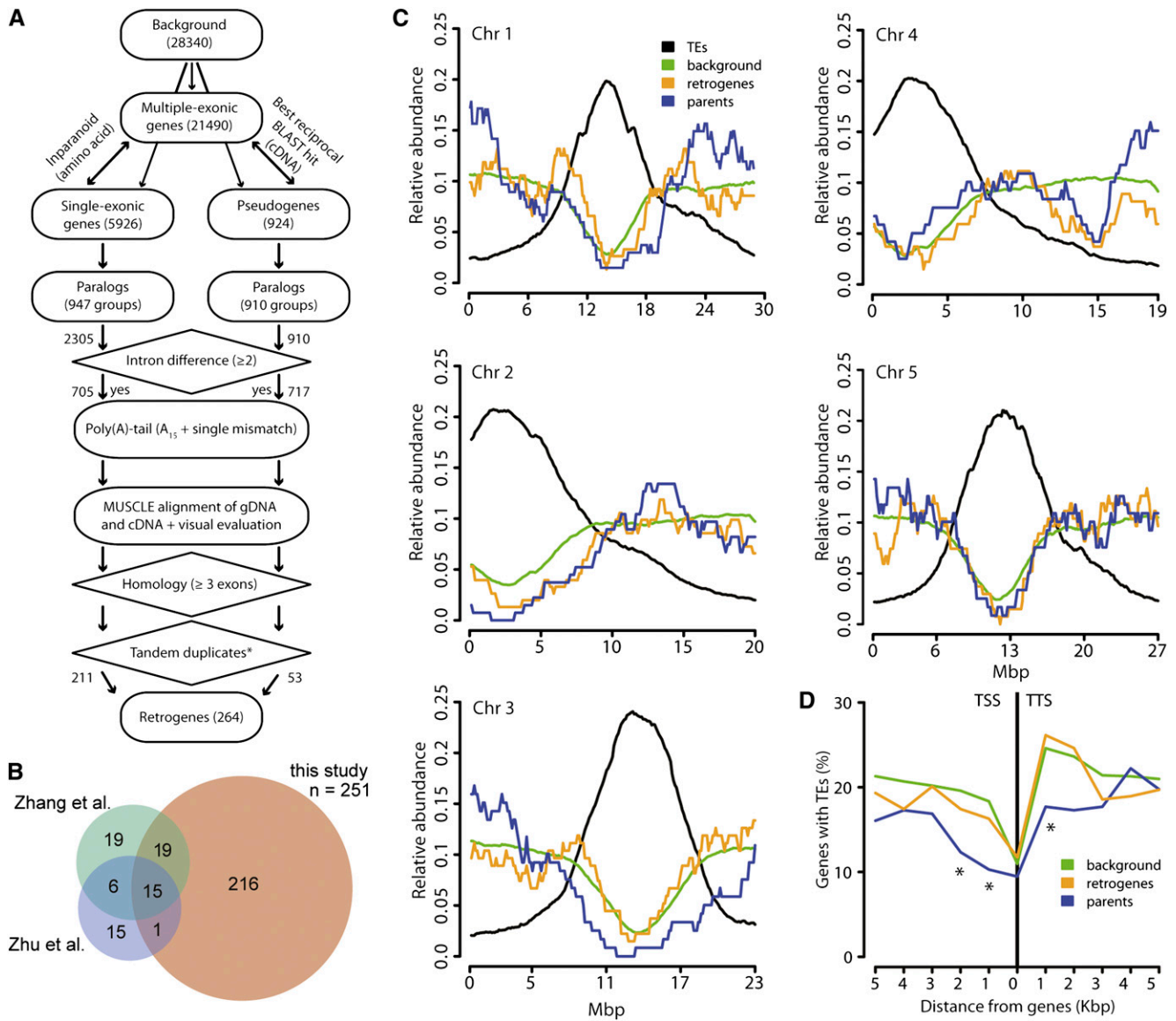
**Figure 1.** Retrogene Identification and Genomic Features.

**(A)** Schematic representation of the retrogene identification method developed for our study.

**(B)** Venn diagrams indicating the numbers of retrogenes identified in three *Arabidopsis* genome-wide searches (Zhang et al., 2005; Zhu et al., 2009; this study). Note that the Venn diagrams do not include the disputable retrogenes listed in Supplemental Data Set 2.

**(C)** Relative abundance (*y* axis) of TEs (black), genes and pseudogenes (background; green), retrogenes (red), and parents (blue) over the five *Arabidopsis* chromosomes (*x* axis).

**(D)** Percentage of genes containing TEs (*y* axis) in 1-kb intervals from the gene TSS and TTS for all protein-coding genes (background; green), retrogenes (red), and parents (blue). Significant differences ($P < 0.05$) in the $\chi^2$ test relative to background are indicated by asterisks.

enriched for close-lying TEs compared with the genomic average, but parents are depleted of TEs in both upstream and downstream intergenic regions.

Hence, the *Arabidopsis* genome contains at least 291 retrogenes located predominantly in gene-rich chromosomal regions. About 10% of the parents gave rise to multiple retrogenes, and ~4.3% of the retrogenes underwent a second retroposition.

**Retrogenes Are Derived from Highly Transcribed Parental Genes and Are Transcribed Preferentially by Novel Promoters**

We took advantage of the comprehensive retrogene list assembled in our study and explored the patterns of retrogene transcription in *Arabidopsis*. The mRNA accumulation was analyzed using microarray data from the 49 *Arabidopsis* developmental

stages assembled by the AtGenExpress consortium (Schmid et al., 2005) and validated for selected tissues by RNA sequencing (Loraine et al., 2013). In total, 209 retrogenes and 245 parents are present on the ATH1 cDNA microarray (Supplemental Data Sets 4 and 5). To compare the effects of RNA- and DNA-based duplications, we also analyzed the set of 3088 *Arabidopsis* DNA duplicated genes (Blanc and Wolfe, 2004). Plotting the mean $\log_2$ robust multiarray averaging (gcRMA; Irizarry et al., 2003) values of all ATH1 probe sets ($n = 22,746$) revealed a double-peak distribution, with the left peak representing genes with poor mRNA levels and/or background signals (Supplemental Figure 2 and Supplemental Data Set 5). The gcRMA values of some retrogenes and parents overlapped with this region and suggested that some of the candidates may not be transcribed in any of the 49 stages. Therefore, we kept only the genes with gcRMA values of 5 or higher in at least one developmental stage (transcribed

genes). In total, 89.4% ($n = 20,398$) of all genes, 85.2% ($n = 178$) of retrogenes, 94.7% ($n = 232$) of parents, and 99.3% ($n = 3067$) of DNA duplicated genes passed these criteria (Figure 2A; Supplemental Data Set 5). This shows that the majority of *Arabidopsis* retrogenes are transcribed in at least some developmental stages, and their mean gcRMA values did not differ significantly from the genome-wide gene set (MWW test, $P = 0.48$; Figure 2A). The parents were significantly enriched for highly transcribed genes relative to both retrogenes and the whole-genome set (MWW test, $P = 7.64 \times 10^{-06}$ and $P = 1.86 \times 10^{-11}$, respectively; Figure 2A). Similarly, DNA duplicated genes were strongly transcribed and therefore similar to parents, but they were strongly different from retrogenes (MWW test, $P = 0.16$ and $P = 1.56 \times 10^{-10}$, respectively). To reveal the transcription relationships between individual retrogene/parent pairs, we compared their developmental stage–specific gcRMA ratios with the transcription of
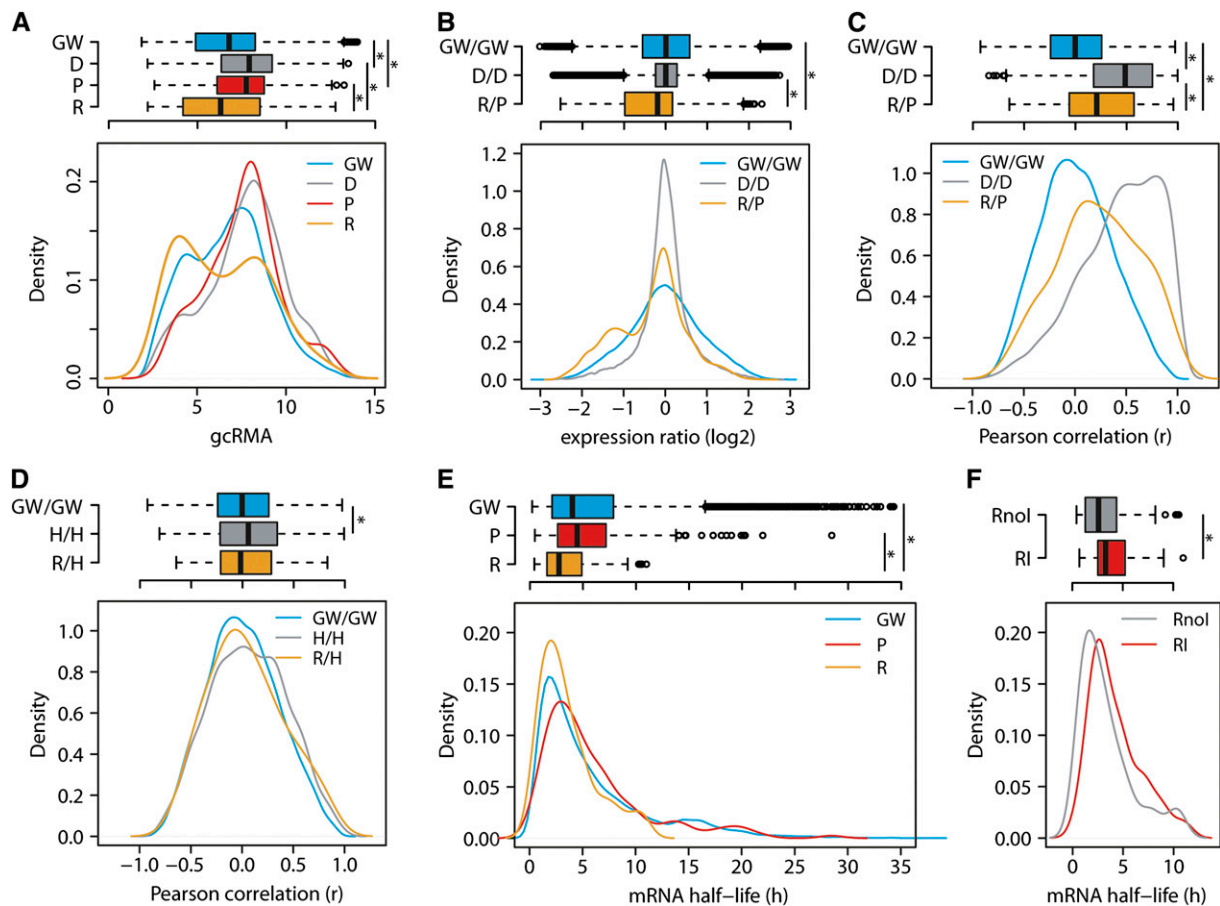


**Figure 2.** Retrogenes Are Driven by Novel Promoters and Have Reduced Transcript Stability.

**(A)** Box and density plots of gcRMA values for genome-wide genes (GW), DNA duplicated genes (D), parents (P), and retrogenes (R) over the 49 *Arabidopsis* developmental stages.

**(B)** $\log_2$ transcription ratios of the random genome-wide gene pairs (GW/GW), DNA duplicated pairs (D/D), and retrogene/parent pairs (R/P).

**(C)** and **(D)** Pearson correlation of gene cotranscription between random genome-wide gene pairs, DNA duplicated pairs, retrogene/parent pairs, genome-wide head-to-head oriented genes (H/H), and retrogene head-to-head oriented neighboring genes (R/H) in 49 developmental stages.

**(E)** and **(F)** mRNA half-lives of genome-wide genes, parents, retrogenes, intronless retrogenes (Rnol), and intronized retrogenes (RI).

Significance values were calculated using the MWW test for all group combinations within each graph, and significant differences ($P < 0.05$) are indicated by asterisks in box plots. Nonsignificant ($P \geq 0.05$) relationships are not shown.

5000 randomly selected gene pairs and the 1527 DNA duplicated gene pairs (Figure 2B; Supplemental Data Set 6). Transcript accumulation ratios of random pairs and DNA duplicated genes represented a broad and narrow range of normally distributed values (MWW test, P = 0.85). Although many retrogenes have a comparable degree of transcription relative to their parents, there is a specific group of 2- to 3-fold less transcribed retrogenes, making retrogene/parent pairs significantly different from both the random gene set and DNA duplicated genes (MWW test, both comparisons P < 2.2 × $10^{-16}$; Figure 2B). Inspecting the gcRMA values over individual developmental stages for the low-transcribed group revealed that these retrogenes were transcribed above the threshold (gcRMA ≥ 5) in only one or a few tissues, while their parents frequently showed ubiquitous transcription.

A recent study in rice suggested frequent cotranscription between retrogenes and parents in plants (Sakai et al., 2011). Our retrogene identification criteria and the nature of the *Arabidopsis* retrogenes (e.g., an absence of retrogenes residing in the introns of other genes) allowed testing three possible mechanisms of retrogene *cis*-regulatory element origin: (1) carryover of parental promoters, (2) the use of bidirectional promoters, and (3) an acquisition of novel *cis*-regulatory elements. First, we tested whether the *Arabidopsis* retrogenes inherit the parental transcription pattern. We calculated the cotranscription of retrogene/parent pairs as Pearson product–moment correlation coefficients (*r*) across the 49 developmental samples of the AtGenExpress data set. Indeed, cotranscription in the set of retrogene/parent pairs (n = 179) was significantly higher than in the 20,000 randomly selected gene pairs (MWW test, P = 2.30 × $10^{-6}$) (Figure 2C; Supplemental Data Set 4). We calculated the frequencies of genes per 0.1 *r* correlation bins for retrogenes and genome background and used this to calculate the number of highly cotranscribed retrogene/parent pairs. In total, 25% of the retrogene/parent pairs (26 out of 102) were correlated more than random gene pairs. However, the cotranscription of DNA duplicated gene pairs, calculated in the same way, was more prominent (MWW test, P < 2.2 × $10^{-16}$; Figure 2C), and 45.6% of them surpassed the random-pairs background.

Second, we tested the possibility for retrogene transcription by bidirectional promoters of head-to-head oriented neighboring genes (Supplemental Data Set 7). The Pearson correlations of random transcribed gene pairs (n = 20,000) and the genome-wide set of transcribed head-to-head oriented genes (n = 2,087; Supplemental Data Set 8) revealed an infrequent but consistent cotranscription between head-to-head oriented gene pairs (MWW test, P = 2.705 × $10^{-10}$; Figure 2D). This shows that sharing bidirectional *cis*-elements is not common in *Arabidopsis*. Retrogene–head-to-head oriented neighbor pairs (n = 63) displayed an intermediate pattern that was not significantly different from either genome-wide or head-to-head oriented genes (MWW test, both P = 0.60; Figure 2D). Only 2.5% of head-to-head oriented retrogenes had higher correlation than random pairs, illustrating the negligible effect of promoter sharing (Figure 2D).

Most retrocopies are expected to be intronless at the time of integration. However, approximately one-third of retrogenes we found contained introns. This indicated that retrogene intronization has a functional role. We tested whether intronization plays a role in retrogene mRNA stability. First, we compared the mRNA half-life of transcribed retrogenes (n = 100), parents (n = 147), and the genome-wide set of transcribed genes (n = 13,012) included in the publicly available mRNA decay data set (Narsai et al., 2007). The mRNA half-life of the parents and the genome-wide gene set was similar (MWW test, P = 0.21) and significantly longer than that of the retrogene mRNA (MWW test, P = 3.56 × $10^{-5}$ and P = 2.54 × $10^{-5}$, respectively; Figure 2E). Furthermore, mRNA of intron-containing retrogenes (29%) had a slightly but significantly longer half-life compared with that of intronless retrogenes (MWW test, P = 0.04; Figure 2F).

Hence, retrogenes are transcribed more weakly than their parents and the transcription of most of the retrogene/parent pairs is not correlated, due to the acquisition of novel regulatory elements at their integration sites. Retrogene mRNA half-life is increased by intronization.

### *Arabidopsis* Retrogenes Are Transcribed in Male Gametes

To analyze the developmental regulation of *Arabidopsis* retrogene transcription, we plotted the mean gcRMA values of genome-wide, parent, and retrogene sets for each of the 49 analyzed developmental stages (Figure 3A). The average mRNA level of parents was higher than that of retrogenes and the genome-wide gene set in all stages. The mean transcription per group was relatively constant, except for pollen, where there was a dip in transcription in the parents and the genome-wide set that contrasted with a peak of retrogene transcription (Figure 3A). To identify relationships between developmental stages and retrogenes, we hierarchically clustered both groups and expressed the result as a heat map of the retrogene transcription z-scores (Figure 3B). This separated stamen and pollen from the rest of the tissues. The highest frequency of retrogenes with positive z-scores (z > 0) was then found in pollen and seeds (62 and 63%, respectively; Figure 3C). However, with more stringent criteria (z > 1 and z > 3), the pollen peak became more prominent relative to other tissues and corresponded to 50 and 30% of retrogenes, respectively (Figure 3C). This shows that many retrogenes reach their transcription maxima in pollen. The pollen-specific transcription pattern has been confirmed by an analysis of individual cases (Figure 3D; Supplemental Figure 3A) and resembles the pattern of retrogene activation in the testis of insects and mammals (reviewed in Kaessmann, 2010).

However, plotting the transcription quantiles (Supplemental Data Set 9) of retrogene gcRMA revealed that not all retrogenes followed this simple trend and that the retrogenes with a negative z-score (pollen downregulated) were usually derived from the group of developmentally highly transcribed genes (Figure 3E, bottom). Remarkably, this distribution also held true for the genome-wide gene set (Figure 3D, top). The parents and the DNA duplicated genes showed more prominent downregulation of the highly transcribed genes (quantile 4) and less obvious upregulation of lowly transcribed genes (quantile 1), while TEs showed upregulation for all quantiles (Supplemental Figure 3B). Hence, we found a pollen-specific activation of retrogenes that is a part of the global pollen-specific transcriptional reprogramming.
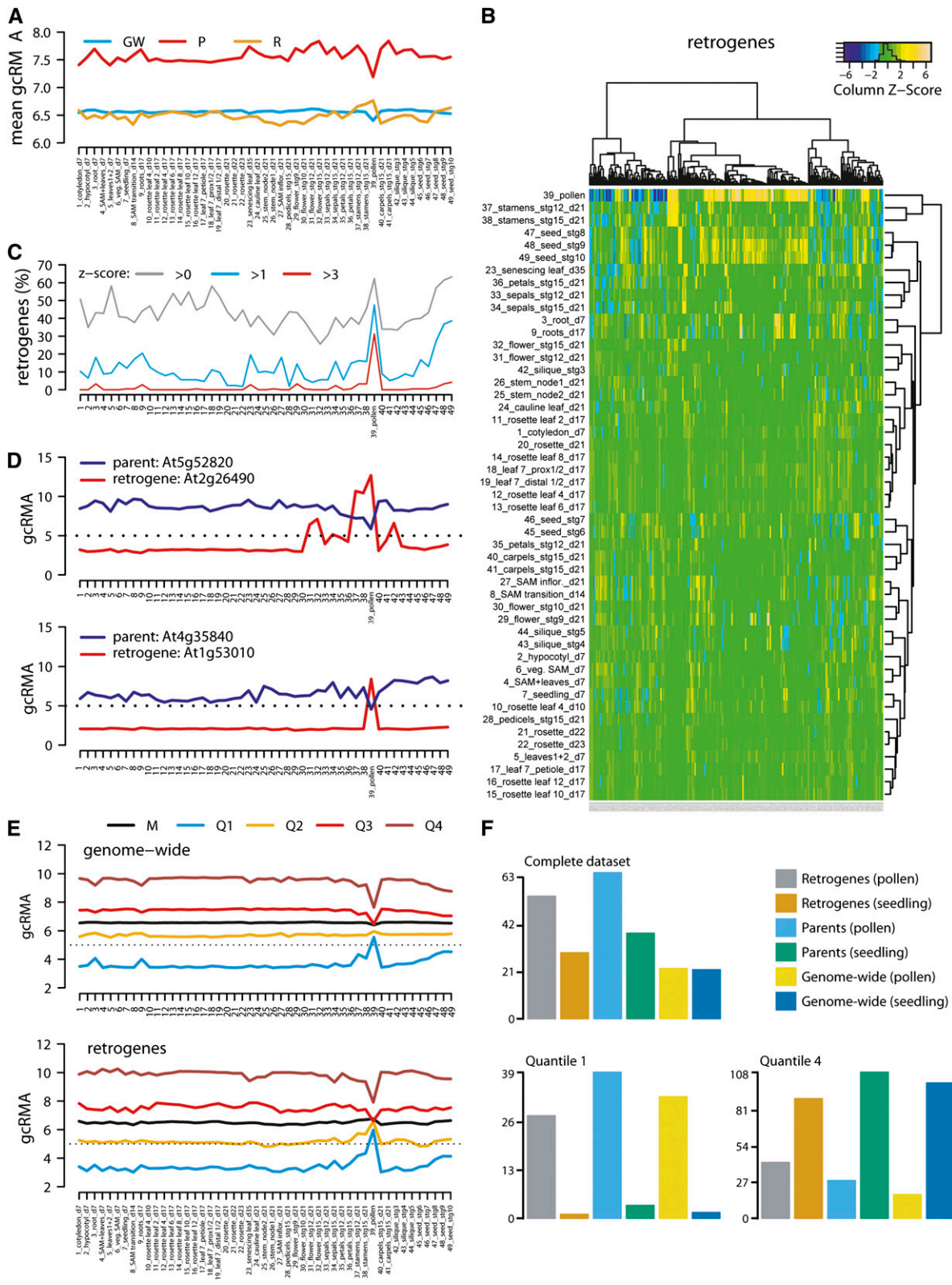
**Figure 3.** Retrogenes Are Transcriptionally Upregulated in Pollen.

Pollen development includes several steps (Honys and Twell, 2003). To find out whether retrogenes are transcribed in specific pollen developmental stages, we compared their transcription in unicellular microspores, bicellular pollen, tricellular pollen, and two highly correlated ($r$ = 0.92) samples of mature pollen grains (Honys and Twell, 2004; Schmid et al., 2005). This revealed a continuous increase of mean retrogene transcription throughout pollen development that contrasted with the downregulation of parental genes in tricellular pollen and mature pollen grains (Supplemental Figure 3C). Next, we asked whether there is an enrichment for retrogene transcripts in vegetative and sperm cells (Honys and Twell, 2003). We used TEs as the control for vegetative cell–specific transcription based on the recently proposed model (Slotkin et al., 2009). Although we observed strong TE upregulation in pollen relative to leaves (MWW test, P < 2.2 × $10^{-16}$), there was a significantly higher amount of TE transcripts in sperm cells relative to the entire pollen (MWW test, P = 0.013; Supplemental Figure 3D). This indicates that there is a higher amount of TE transcripts in both pollen cell types. The parents were significantly more highly transcribed in sperm cells relative to seedlings (MWW test, P = 0.001) and were underrepresented for the lowly transcribed genes in this tissue relative to entire pollen (Figure 3E). Therefore, retrogene parents are transcribed preferentially in sperm cells. The median of retrogene transcription was higher than that of TEs and increased in both pollen samples relative to seedlings, but only the entire pollen differed significantly (MWW test, P = 0.008; Figure 3E). In combination with pollen developmental stage data, this shows that retrogenes are transcribed in both pollen cell types.

In order to validate our results by independent experiment, we tested whether our findings hold true in data sets generated by RNA sequencing. Gene transcription in mature pollen grains was compared with that in seedling tissues (Loraine et al., 2013). Plotting the mean reads per kilobase per million reads values for the entire set, quantile 1 (lowest transcribed), and quantile 4 (highest transcribed) of all genes, retrogenes, and parents confirmed the microarray data (Figures 3A, 3E, and 3F; Supplemental Figure 3B). The only exception was a higher transcription of parents in pollen relative to seedlings in RNA sequencing (mean and quantile 1 samples; absent in the quantile 4 sample), while the opposite results were obtained using microarrays (Figures 3A and 3F). This difference is due to the higher sensitivity of RNA sequencing technology to quantify transcripts from lowly transcribed genes (Mooney et al., 2013; Zhao et al., 2014). This partially applies also to retrogenes, as the upregulation in pollen versus seedling is more

pronounced in RNA sequencing compared with microarrays (Figure 3F).

From this, we conclude that retrogene activation starts prior to pollen maturation and later occurs in both terminal pollen cell types.

## Retrogenes Are Deficient for Transcription-Permissive Chromatin Marks in Leaf Tissues

Analysis of transcription quantiles suggests that global transcriptional changes in pollen have a major effect on retrogene transcription. This may be achieved by a global chromatin reprogramming (Kaessmann et al., 2009). Therefore, we calculated $\log_2$-fold transcription changes between pollen and 21-d-old rosettes (ATGE_73/ATGE_22; Schmid et al., 2005) and correlated those with transcriptional changes induced by chromatin mutants (mutant rosettes/wild-type rosettes). Five groups were compared (Supplemental Data Set 4): all genes ($n$ = 22,746), pollen upregulated genes ($n$ = 5171), leaf upregulated genes ($n$ = 6057), pollen upregulated retrogenes ($n$ = 51), and leaf upregulated retrogenes ($n$ = 53). Tissue upregulated genes were defined as having $\log_2$-fold change $\geq$ 1 in one versus the other tissue. First, we estimated the effects of the transposon-silencing machinery by testing mutants for *DECREASED DNA METHYLATION1* (*DDM1*), *KRYPTONITE* (*KYP*), and *HISTONE DEACETYLASE6* (*HDA6*) (Baubec et al., 2010; Inagaki et al., 2010; Popova et al., 2013), which lead to a loss of DNA methylation, a loss of H3K9me2, and a gain of histone acetylation at heterochromatic loci, respectively. There was no clear correlation (maximum $r$ = 0.040) between transcription in pollen relative to leaves and transcriptional changes induced by *ddm1*, *kyp*, and *hda6* for all tested groups (Supplemental Figures 4A to 4C). This demonstrates that TE silencing components do not determine the global gene transcription pattern in pollen or affect retrogenes. Next, we tested the effects of the H3K27me3 mark by analyzing mutants of the Polycomb group repressive complex factors *CURLY LEAF* (*CLF*) and *SWINGER* (*SWN*), which have been shown to regulate transcription during development (Farrona et al., 2011; Lafos et al., 2011). The correlation between *clf* and *swn* single mutants with pollen-specific transcriptional changes was low ($r$ < 0.20; Supplemental Figures 4D and 4E). Because CLF and SWN are partially functionally redundant (Lafos et al., 2011), we tested for effects in the *clf swn* double mutant. The correlation between pollen and *clf swn* transcription profiles for the set of all genes was higher ($r$ = 0.277) than for the *clf* and *swn* single mutants (Figure 4A; Supplemental Figures 4D

---

**Figure 3.** (continued).

**(A)** Mean gcRMA values for genome-wide genes (GW), parents (P), and retrogenes (R) at each of the 49 *Arabidopsis* developmental stages.

**(B)** Hierarchically clustered heat map of retrogene z-scores (*y* axis) and developmental stages (*x* axis).

**(C)** The frequency of retrogenes with row z-scores in **(B)** >0, >1, and >3 in individual developmental stages.

**(D)** Examples of retrogenes and parents showing tissue-specific and ubiquitous transcription, respectively, with major transcription changes in pollen (stage 39).

**(E)** Developmental gcRMA values for the genome-wide set of genes and retrogenes. Transcription is shown for mean (M) and transcription quantiles: lowly transcribed/quantile 1 (Q1), mid-lowly transcribed/quantile 2 (Q2), mid-highly transcribed/quantile 3 (Q3), and highly transcribed/quantile 4 (Q4).

**(F)** Mean RNA sequencing reads per kilobase per million reads (RPKM) values for all genes (genome-wide), parents, and retrogenes in vegetative rosettes and pollen as complete data sets, quantile 1 (lowly transcribed genes), and quantile 4 (highly transcribed genes).

and 4E). Surprisingly, the high correlation was mainly due to leaf upregulated genes and retrogenes ($r$ = 0.469 and 0.364, respectively) that were coordinately downregulated in both pollen and *clf swn* (Figure 4A). By contrast, pollen upregulated genes showed generally uncorrelated transcription with *clf swn* ($r$ = −0.047). To further test the connection with H3K27me3 changes, we analyzed transcription in a mutant for *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*), another key gene of the Polycomb repressive complex (Bouyer et al., 2011). Although the correlations between *fie* and pollen transcription profiles

were weaker ($r$ = 0.186, 0.366, and 0.268 for all genes, leaf upregulated genes, and retrogenes, respectively; Figure 4B), they perfectly recapitulated trends observed in the comparison between *clf swn* and pollen. Hence, loss of key components of the Polycomb repressive complex correlates with pollen-specific gene downregulation of leaf transcribed genes but does not explain pollen-specific gene upregulation.

Therefore, we used publicly available chromatin data from young *Arabidopsis* leaves (Roudier et al., 2011) to test which chromatin modification(s) is associated with retrogenes and
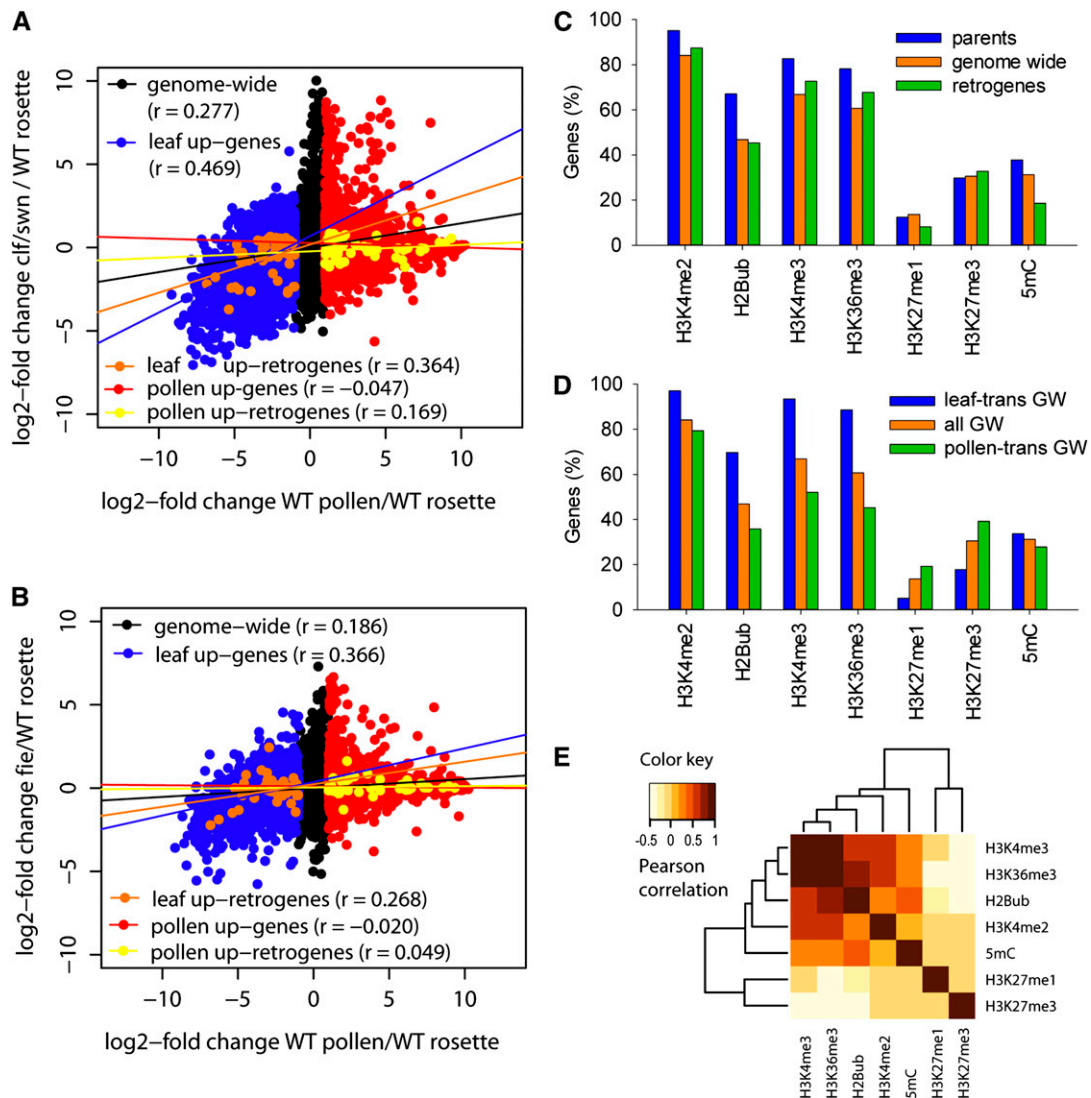


**Figure 4.** Chromatin Regulation of Pollen-Specific Gene Transcription.

**(A)** and **(B)** Dot plots of log$_2$-fold changes in wild-type pollen/rosettes (*x* axis) and *clf swn* double mutants (*y* axis) **(A)** or *fie*/wild-type rosettes (*y* axis) **(B)**. Specific gene sets were superimposed on the genome-wide set in different colors. The lines indicate transcription correlation (*r*) between the *x* and *y* axes for specific gene sets. The *r* values are given in parentheses.

**(C)** The frequency of seven chromatin modifications at gene-coding sequences for all genes, parents, and retrogenes in young leaf tissues.

**(D)** The same as **(C)** for all genes (all GW), leaf transcribed genes (leaf-trans GW), and pollen transcribed genes (pollen-trans GW).

**(E)** Hierarchical clustering and heat map of Pearson correlation values of colocalization between seven chromatin modifications for all *Arabidopsis* genes.
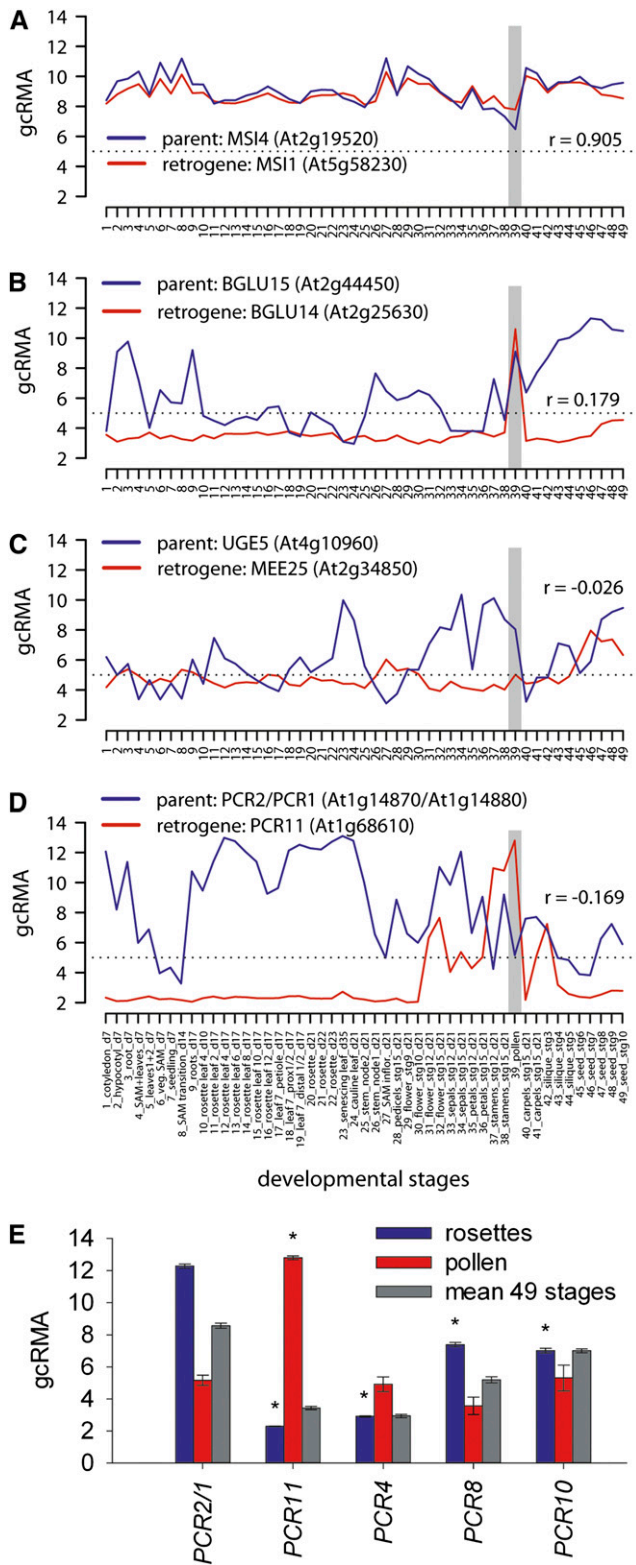
## Figure 5 (A–E)

**A** — gcRMA vs developmental stages
parent: MSI4 (At2g19520)
retrogene: MSI1 (At5g58230)
r = 0.905

**B** — gcRMA vs developmental stages
parent: BGLU15 (At2g44450)
retrogene: BGLU14 (At2g25630)
r = 0.179

**C** — gcRMA vs developmental stages
parent: UGE5 (At4g10960)
retrogene: MEE25 (At2g34850)
r = -0.026

**D** — gcRMA vs developmental stages
parent: PCR2/PCR1 (At1g14870/At1g14880)
retrogene: PCR11 (At1g68610)
r = -0.169

developmental stages

Developmental stage labels:
1_cotyledon_d7, 2_hypocotyl_d7, 3_root_d7, 4_SAM-Leaves_d7, 5_leaves1+2_d7, 6_veg_SAM_d7, 7_seedling_d14, 8_SAM transition_d14, 9_roots_d21, 10_rosette leaf 4_d10, 11_rosette leaf 2_d17, 12_rosette leaf 4_d17, 13_rosette leaf 6_d17, 14_rosette leaf 8_d17, 15_rosette leaf 10_d17, 16_rosette leaf 12_d17, 17_leaf 7_petiole_d17, 18_leaf 7_prox1/2_d17, 19_leaf 7_distal 1/2_d17, 20_rosette_d21, 21_rosette_d22, 22_rosette_d23, 23_senescing leaf_d35, 24_cauline leaf_d21, 25_stem_node1_d21, 26_stem_node2_d21, 27_SAM inflor._d21, 28_pedicels_stg15_d21, 29_flower_stg9_d21, 30_flower_stg10/11_d21, 31_flower_stg12_d21, 32_flower_stg15_d21, 33_sepals_stg15_d21, 34_sepals_stg12_d21, 35_petals_stg15_d21, 36_petals_stg12_d21, 37_stamens_stg15_d21, 38_stamens_stg12_d21, 39_pollen, 40_carpels_stg15_d21, 41_carpels_stg12_d21, 42_silique_stg3, 43_silique_stg4, 44_silique_stg5, 45_seed_stg5, 46_seed_stg6, 47_seed_stg8, 48_seed_stg9, 49_seed_stg10

**E** — gcRMA: rosettes, pollen, mean 49 stages
PCR2/1, PCR11, PCR4, PCR8, PCR10

**Figure 5.** Gain of Pollen-Specific Transcription by the *PCR11* Retrogene.

**(A)** to **(D)** Developmental gcRMA transcription profiles of retrogenes associated with pollen growth and development and their parents. Pollen stage is highlighted by the vertical gray bars.
**(E)** gcRMA transcription values of *PCR* family genes in rosettes, pollen, and mean of 49 developmental stages and tissues. *PCR2* and *PCR1* correspond to a single microarray element and therefore are shown together. Transcription values were compared with *PCR2/PCR1* transcription in the same tissue, and statistically significant differences (P < 0.05) in *t* tests are labeled with asterisks. Error bars denote SD of three biological replicates.

pollen upregulated genes in somatic tissues. We extracted information on chromatin marks for every gene and compared the full sets of retrogenes, parents, and all genes (Figure 4C). In accordance with high and ubiquitous transcription, the parents were enriched for the permissive chromatin marks histone H3 Lysine 4 dimethylation and histone H3 Lysine 4 trimethylation (H3K4me3), histone H3 Lysine 36 trimethylation (H3K36me3), and histone H2B ubiquitination (H2Bub), followed by retrogenes and the genome-wide set. None of these groups was enriched for the repressive histone H3 Lysine 27 modifications. The enrichment for gene body DNA methylation in highly expressed genes is consistent with the currently proposed function of this modification (Coleman-Derr and Zilberman, 2012). Next, we compared previously defined groups of pollen upregulated and leaf upregulated genes (Supplemental Data Set 10). The pattern of distribution of chromatin marks for each individual group (retrogenes, parents, and all genes) was relatively similar (Figure 4D; Supplemental Figures 5A and 5B). There were no changes in gene body DNA methylation. While the histone H3 Lysine 27 modifications were enriched in pollen upregulated genes of the genome-wide set, this mark does not seem to play a major role in the somatic silencing of pollen upregulated retrogenes (Supplemental Figure 5A). By contrast, all analyzed transcription-permissive marks (histone H3 Lysine 4 dimethylation, H3K4me3, H2Bub, and H3K36me3) were underrepresented in pollen upregulated genes in leaf tissues (Figure 4D; Supplemental Figures 5A and 5B). The presence or absence of these marks was strongly correlated in pairwise comparisons of individual modifications (Figure 4E; Supplemental Figure 5C).

This suggests that in leaf tissues, retrogenes and other pollen upregulated genes are depleted of permissive chromatin marks without enrichment for repressive marks. By contrast, leaf upregulated genes are downregulated in pollen by a mechanism involving the Polycomb repressive complex components *CLF*, *SWN*, and *FIE*.

### Gain of Transcription Factor Binding Sites Facilitates *PCR11* Retrogene Sperm-Specific Transcription

Next, we tested whether the gamete-specific transcription of retrogenes has evolved into gamete-specific developmental functions. Five retrogenes found in our screen, *MULTICOPY SUPRESSOR OF IRA1* (*MSI1*), *PLANT CADMIUM RESISTANCE11* (*PCR11*), *BETA-GLUCOSIDASE14* (*BGLU14*), *MATERNAL EFFECT EMBRYO ARREST25* (*MEE25*), and *PEROXIDASE*, are associated with pollen development, sperm cell differentiation, pollen tube

growth, and development (TAIR10). To investigate the relationship between the transcription of these retrogenes and their parents, we plotted their mean developmental gcRMA values and calculated transcription Pearson correlations (Figures 5A to 5D; Supplemental Data Set 4). The parental gene of *PEROXIDASE* was not included on the ATH1 array; therefore, we did not continue its analysis. The transcription of *MSI1* was strongly correlated ($r = 0.905$) with its parent *MSI4*, and both were ubiquitously transcribed throughout development (Figure 5A). *BGLU14* and its parent *BGLU15* were both upregulated in pollen (Figure 5B). The *MEE25* retrogene was transcribed at low levels throughout development, and higher transcription was found only in embryonic tissues (Figure 5C). However, its parent, At4g10960, was transcribed mainly in floral tissues, seeds, and pollen, where it greatly surpassed *MEE25* transcription. Hence, these three retrogenes did not provide evidence for the development of parent-independent pollen-specific transcription. In contrast, *PCR11* was transcribed at low levels almost throughout the entire process of development but was activated in floral tissues, stamen, and pollen. This pattern was opposite to that of its parent, *PCR2*, which was active mainly in the photosynthetically active tissues and downregulated in stamen and pollen (Figure 5D). *PCR11* has been shown to be transcribed specifically in pollen sperm cells by the MYB transcription factor DUO1 (Borg et al., 2011). Therefore, we compared the promoter regions of *PCR11* and *PCR2* and looked for previously described DUO1 binding motifs (Borg et al., 2011). There are three binding regions in the 500-bp region upstream of the *PCR11* TSSs (TAACCGTC at −47 to −54 bp and AAACCG at −153 to −158 and −452 to −457 bp). However, only a single DUO1 binding motif (AAACCGT at −100 to −106 bp from the TSS) is found in the promoter of *PCR2*. To test whether this represents a gain of function in *PCR11* or a loss of function in *PCR2*, we compared the promoter regions of several other *PCR* family members representing both the *PCR2* clade (*PCR1* and *PCR3*) and the outgroups (*PCR4*, *PCR8*, and *PCR10*) (Song et al., 2010). None of these genes contained a single DUO1 binding motif in the 500-bp region upstream of the TSS. Furthermore, comparing their transcript levels revealed that only *PCR11* is significantly upregulated in pollen relative to *PCR2* (Figure 5E).

Next, we tested these results in an independent experiment by analyzing retrogene and parent transcription in *Arabidopsis* lines carrying somatically inducible DUO1 (Borg et al., 2011). Upon 6, 12, and 24 h of DUO1 induction, we observed 36, 131, and 125 significantly upregulated genes and 47, 124, and 121 significantly downregulated genes, respectively. The number of upregulated and downregulated retrogenes (2 and 1, respectively) was small (Supplemental Data Set 11), showing that DUO1 regulates the transcription of only a few specific retrogenes. Importantly, the set of significantly upregulated retrogenes included the *PCR11* retrogene (log2-fold changes in 6, 12, and 24 h: 0.26, 2.21, and 4.03; $t$ test P values: 0.010, 3.3 × $10^{-5}$, and 5.4 × $10^{-5}$, respectively). This has been reflected by significant downregulation of its parent *PCR2* in two out of three experimental points (log2-fold changes in 6, 12, and 24 h: −1.18, −1.60, and −0.74; $t$ test P values: 0.003, 0.007, and 0.19, respectively). Therefore, we conclude that the *PCR11* retrogene gained sperm cell–specific DUO1-dependent transcription relative to its parent *PCR2*.

## DISCUSSION

### Multiple and Repeated Retropositions in *Arabidopsis*

We found 251 retrogenes in *Arabidopsis*, 216 of which are newly identified. The limited overlap of our set with the previous *Arabidopsis* retrogene lists was most likely due to partly different search criteria and thresholds of individual methods (Zhang et al., 2005; Zhu et al., 2009). We detected ~50% of the retrogenes found in the study of Zhang et al. (2005). A specific subset of the remaining retrogenes was not accepted by our method, owing to different thresholds for selection or a lack of positive evidence for retroposition, such as missing information on the parental gene or insufficient difference in intron number (Supplemental Data Set 2). The smaller (43.2%) overlap with the set identified by Zhu et al. (2009) is due to their use of very specific criteria to identify chimeric retrogenes. These criteria apparently hamper the identification of structurally simple retrogenes; conversely, our method does not allow the identification of chimeric retrogenes. The higher number of retrogenes detected with our analysis is due to several factors: (1) search among *Arabidopsis* pseudogenes, (2) allowing intronized retrogenes, and (3) accepting multiple retrocopies derived from a single parent (applied also in Zhang et al., 2005). Although we increased the number of retrogenes in *Arabidopsis* 3-fold, our selection criteria were conservative and the current number is most likely an underestimate based on two facts. First, we omitted several hundred candidates that had at least one paralog within the *Arabidopsis* genome but did not show evidence of retroposition [i.e. did not differ by two or more introns or did not have a poly(A) tail]. Second, none of the plant genome-wide retrogene screens detected retrogenes of the Ser-Glu-Thr domain protein group (Zhang et al., 2005; Zhu et al., 2009; this study), which were identified in studies focusing specifically on the evolution of this gene family (Baumbusch et al., 2001; Zhu et al., 2011). Hence, 1% of *Arabidopsis* genes estimated to be retrogenes is most likely an underestimation.

Although we have found 3-fold more retrogenes in *Arabidopsis* than were previously found in rice (Sakai et al., 2011), the number of conservatively estimated retrogenes per plant genome is much smaller compared with metazoans (e.g., 19.1% in human) (Marques et al., 2005; Pennisi, 2012). This difference may have multiple reasons. Since most of the retrogenes are identified based on intron loss, greater intron numbers in parents would simplify retrogene identification. This may partially explain the difference between the genomes of *Arabidopsis* and human, which have average numbers of 4.2 and 7.8 introns per gene, respectively (Arabidopsis Genome Initiative, 2000; Sakharkar et al., 2004). Another possibility, which is not mutually exclusive, builds on the scarcity of WGDs in many groups of higher animals compared with plants (Gregory and Mable, 2005). This may favor local gene duplication mechanisms, including retroposition, in metazoa versus plants. Finally, the higher activity of *LONG INTERSPERSED ELEMENT* (*LINE*) element reverse transcriptases may be responsible for an increased retroposition rate in animals (Beck et al., 2010).

In contrast with animals, where 82% of retrocopies contain premature stop codons (Marques et al., 2005), only 17.4% of

*Arabidopsis* retrogenes are annotated as pseudogenes. This suggests a higher retrogene success rate in plants relative to the total number of retrocopies. Further support comes from our observation that several retrogenes served as parents and produced secondary retrocopies. Therefore, retroposition contributes to functional plant genome evolution.

One of the unresolved questions in retrogene biology is how transcripts are selected for retroposition. Although retroposition in animals has been associated with *LINE* element amplification machinery, this link has not been firmly proven in plants (Ohshima, 2013). We describe 22 parents that produced up to seven retrogenes each, which suggests one or more common features or a signal for retroposition in *Arabidopsis*. Additional support comes from the 13 cases where a repeated retroposition has been found. Since retrotransposon reverse transcriptases favor specific sequences in combination with transcript folding (Ohshima, 2013), it is possible that such structures exist also in transcripts of some protein genes. Similar to other plant and animal studies (Marques et al., 2005; Potrzebowski et al., 2008; Sakai et al., 2011), we have confirmed that parents are generally strongly and ubiquitously transcribed, indicating that higher amounts of transcript may increase the probability of retroposition. Although produced by the retrotransposon amplification machinery, they are located in gene-rich chromosome arms in *Arabidopsis* and thus fundamentally differ in their genomic distribution from repetitive elements. This also holds true for their upstream and downstream intergenic regions that are not enriched for repetitive DNA.

### *Arabidopsis* Retrogenes Are Transcribed via Newly Acquired Promoters

One of the major limitations to the establishment of retrogenes as functional genes is the loss of *cis*-regulatory sequences (Kaessmann et al., 2009). Hence, we analyzed the retrogene transcription in *Arabidopsis* using genome-wide transcription data of 49 different *Arabidopsis* developmental stages by microarrays. In agreement with the observations in rice (Sakai et al., 2011), we found that retrogenes are transcribed less compared with their parents. However, retrogene transcription resembles the whole-genome average, suggesting that they are not "dead on arrival" in *Arabidopsis*. The parents are mostly recruited from highly and ubiquitously transcribed genes, indirectly supporting the hypothesis that transcript abundance is an important prerequisite for retroposition.

In human, it has been shown that retrogenes and parents may share promoter sequences, implying a carryover of the parental promoter by retroposition of transcripts from an upstream TSS (Okamura and Nakai, 2008). Furthermore, a recent study in rice revealed a number of retrogene/parent pairs with positively correlated transcription profiles among seven developmental stages (Sakai et al., 2011). However, this analysis did not include correction for the cotranscription of random gene pairs (Sakai et al., 2011); therefore, the extent of correlation may be partially overestimated. Our data show that ~25% of retrogene/parent pairs and 3% of retrogene head-to-head oriented neighboring genes are cotranscribed beyond the genome background in *Arabidopsis*. Hence, rice and *Arabidopsis* data support the

mechanism of *cis*-regulatory element carryover in plants. However, DNA sequence analysis of parent and retrogene promoters did not reveal significant homology in rice (Sakai et al., 2011). Therefore, it remains unclear whether retrogenes retropose including parental upstream regulatory sequences that mutate rapidly afterward or they carry cryptic exonic regulatory sequences. In *Arabidopsis*, the majority (72%) of retrogenes are transcribed in a pattern that is not correlated to that of parents and neighboring genes, suggesting the acquisition of novel *cis*-regulatory elements in most cases. Currently, it is unknown whether this pattern is the result of postintegration selection or whether the compact *Arabidopsis* genome offers a sufficient density of cryptic promoters.

Previous studies in *Arabidopsis* showed that transcripts of single-exon genes are relatively short-lived (Narsai et al., 2007). We observed that this also holds true for single-exon retrogenes and that retrogene intronization significantly increases their mRNA half-life. Hence, intron retention in the retrogene parent mRNAs and/or retrogene neointronization may help establish retrogenes as mature genes.

### Retrogenes Are Preferentially Upregulated in Pollen

The separation of gametes from somatic cells is very much delayed in plants compared with animals. Therefore, somatic retroposition events in the shoot apical meristems may also be transmitted to the next generations. Thus, we tested for tissue-specific transcription of retrogenes in *Arabidopsis* using a developmental transcription data series (Schmid et al., 2005) and validated these findings using RNA sequencing data sets (Loraine et al., 2013). Surprisingly, this revealed that retrogenes are overtranscribed in pollen while overall transcription was not increased at this stage. However, this pattern was not uniform for the whole group, as lowly transcribed retrogenes became upregulated in pollen while highly transcribed ones were downregulated. In addition, the set of all *Arabidopsis* genes showed a similar trend. Hence, this transcription pattern is not restricted to retrogenes. More likely, many retrogenes are part of global cellular reprogramming in male gametes. So far, chromatin changes in male gametes have been associated mainly with DNA methylation changes (Slotkin et al., 2009; Ibarra et al., 2012), but there is emerging evidence that histone modifications may also contribute to pollen-specific gene reprogramming (Hoffmann and Palmgren, 2013). In order to identify possible causes of the observed pollen-specific transcription, we explored available data on tissue- and mutant-specific transcription and the distribution of chromatin modifications. By comparing transcriptional profiles of pollen and mutants defective in transcriptional gene silencing, we excluded the loss of DNA methylation and H3K9me2 or heterochromatin-specific histone hyperacetylation as the factors leading to global transcription changes in pollen. The analysis of chromatin profiles in leaves revealed that pollen upregulated genes (and retrogenes) are depleted of the transcription-permissive marks (H2Bub, H3K4me3, and H3K36me3) in these tissues. Recently, it was reported that pollen-specific genes are regulated by histone H3 Lysine 27 methylation in *Arabidopsis* (Hoffmann and Palmgren, 2013), but this trend was much less pronounced in our data set. This is due to

the different selection criteria of candidate genes in both studies. Our set of pollen upregulated genes ($n$ = 5171) included the entire (99.1%) set of pollen-specific genes ($n$ = 584; Hoffmann and Palmgren, 2013). This is most likely masking the enrichment for histone H3 Lysine 27 methylation modifications of a specific subset of pollen-transcribed genes in leaves. However, it has to be noted that H3K27me3 modification may regulate pollen-specific transcription indirectly, as suggested by our transcription analysis of the *clf swn* and *fie* mutants. This also holds true for the group of pollen-specific genes associated with histone H3 Lysine 27 monomethylation and H3K27me3 in leaf tissues (Hoffmann and Palmgren, 2013), as only a few of those genes are upregulated in *clf swn* (Supplemental Figure 5D). Unexpectedly, we found correlated downregulation of similar sets of genes (and retrogenes) in pollen and leaves of *clf swn* and *fie* ($r$ = 0.462 and 0.366, respectively). Gene downregulation in response to the loss of a repressive mark is counterintuitive and suggests that the effect is indirect and may be achieved by the activation of specific H3K27me3-regulated suppressors such as microRNAs (Lafos et al., 2011). Based on this, we propose that it is most likely a temporary absence of permissive marks (without strong enrichment for repressive marks) that causes the upregulation of specific genes in pollen relative to somatic tissues.

Pollen-specific transcription of *Arabidopsis* retrogenes was unanticipated and is analogous to retrogene transcription in animal spermatocytes (Marques et al., 2005; Vinckenbosch et al., 2006; Bai et al., 2008). Although the molecular nature of this specific transcription is so far unknown, two explanatory models have been proposed in animals (Kaessmann et al., 2009). The first suggests sperm-specific retroposition and integration into open (and thus more likely to be transcribed) chromatin that allows transcription and perpetuates this behavior. However, our data do not support this model in two respects. First, integration into active chromatin would most likely be reflected by cotranscription between neighboring genes, which was rare in *Arabidopsis*. Second, we observed many nonretrogene genes with pollen-specific transcription. The second model proposes spermatocyte-specific transcriptional reprogramming by global chromatin changes and transcriptional activation of retrogenes and their subsequent functionalization specific to spermatocytes (Marques et al., 2005; Potrzebowski et al., 2008). In plants, pollen has been identified as the hotspot of chromatin reprogramming (Slotkin et al., 2009; Ibarra et al., 2012; Hoffmann and Palmgren, 2013), and we have shown that pollen upregulated genes are depleted from transcription-permissive chromatin marks in somatic tissues. Furthermore, we found several retrogenes that are associated with pollen growth and development and the *PCR11* retrogene, which is transcribed in pollen, contrary to its parent. This is due to the presence of multiple pollen-specific DUO1 transcription factor binding motifs in its promoter. Hence, our data support the second model and suggest that a small number of retrogenes have developed or retained male gamete–specific functions in *Arabidopsis*.

The activation of many normally lowly transcribed genes and the subsequent downregulation of highly transcribed genes just prior to the onset of the next generation is an intriguing pattern with no known molecular function. However, it seems to be present in both plant and animal lineages and suggests evolutionarily conserved or analogous mechanisms that regulate gene transcription during this critical stage of development.

## METHODS

### Retrogene Identification

The principal steps in retrogene identification in *Arabidopsis thaliana* are given in Figure 1A. First, the paralogy groups between sets of intronless ($n$ = 5923) and intron-containing ($n$ = 21,481) protein-coding genes according to TAIR10 were established using protein homologies in InParanoid 4.1 (Remm et al., 2001). When the paralogy group had multiple intron-containing "inparalogs" with different intron numbers, they were also considered for downstream analysis. Similarly, paralogy groups between pseudogenes ($n$ = 924) and intron-containing protein-coding genes were identified as the best reciprocal BLAST hits using cDNA sequences (Altschul et al., 1990). Accepted retrogene–parent candidate pairs had a minimum homology score of $10^{-10}$ and a minimum difference in intron number of two. Intronless genes were also considered as candidates when differing by only a single intron, if the poly(A) tail was detected within 150 or 250 bp downstream of the retrogene candidate stop codon with or without an annotated 3′ untranslated region, respectively. The poly(A) tail was defined as ≥15 consecutive adenines with a single mismatch. We determined poly(A) tail length as the shortest stretch of adenines present significantly above random (Supplemental Figure 6). Since the absence of introns can be due to a loss of splicing signals (intron retention), the homology of exonic and intronic sequences was validated visually. A retrogene was accepted when a minimum of three consecutive homologous exons, spanning two lost introns, were observed (Edgar, 2004). If multiple parents were predicted for a retrogene, we accepted the candidate with the highest pair-wise alignment score in multiple (cDNA) sequence alignment (Larkin et al., 2007). The protocol was executed with customized bioperl and awk scripts (Stajich et al., 2002).

### Genome-Wide Transcription and mRNA Half-Life Analysis

All microarray analyses were based on publicly available data sets. Throughout the study, we used the following ATH1 cDNA microarrays (Affymetrix): wild-type *Arabidopsis* development data produced by the AtGenExpress consortium (Schmid et al., 2005), *Arabidopsis* pollen development and sperm cell data sets NASCARRAYS-48 (Honys and Twell, 2003, 2004), the *ddm1-12* data set deposited at the Gene Expression Omnibus (GEO) as GSE18977 (Baubec et al., 2010), the *kyp* GEO data set GSE22957 (Inagaki et al., 2010), the *clf*, *swn*, and *clf swn* GEO data set GSE20256, and the *hda6* (*rts1-1*) data set NASCARRAYS-538 (Popova et al., 2013). The raw data were processed and normalized using the robust multiarray averaging method (Irizarry et al., 2003) in R software (www.R-project.org) using Bioconductor (www.bioconductor.org) and the affy package. The *fie* transcription values were retrieved from the GEO data set GSE19851 (Bouyer et al., 2011) as the normalized transcription values. Retrogene and parent probes that corresponded to multiple gene models were excluded from genome-wide analysis. The transcription borderline for transcribed genes (gcRMA ≥ 5) was based on the minimal density of genes between peaks indicating absent or background signals versus high transcription signals (Supplemental Figure 2). The *Arabidopsis* mRNA half-life data and rosette- and pollen-specific RNA sequencing data were extracted from previously published data sets (Narsai et al., 2007; Loraine et al., 2013). Randomized sets of genes or gene pairs were generated, plots drawn, and statistical tests calculated in R. The significance of density distributions was tested using the MWW rank-sum test with correction and cotranscription correlation by the Pearson product–moment correlation coefficient ($r$).

## Chromatin Analysis

Chromatin data of 10-d-old *Arabidopsis* seedlings were retrieved from the publicly available genome-wide atlas of chromatin modifications (Roudier et al., 2011). The frequencies for individual groups were compared. Pearson correlations were calculated in Excel (Microsoft), and heat maps were built in R.

## Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL libraries under the following accession numbers: *MSI1*, At5g58230; *PCR1*, At1g14880; *PCR2*, At1g14870; *PCR3*, At5g35525; *PCR4*, At3g18460; *PCR8*, At1g52200; *PCR10*, At2g40935; *PCR11*, At1g68610; *BGLU14*, At2g25630; *BGLU15*, At2g44450; *MEE25*, At2g34850; parent of *MEE25* retrogene, At4g10960; and *PEROXIDASE*, At4g17690. Other genes listed in the supplemental data sets include Arabidopsis Genome Initiative codes.

## Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure 1.** Examples of Repeated Retroposition in *Arabidopsis*.

**Supplemental Figure 2.** Defining Lowly and Highly Transcribed Genes in *Arabidopsis*.

**Supplemental Figure 3.** *Arabidopsis* Retrogenes Are Transcribed in Pollen.

**Supplemental Figure 4.** Transcription Correlations between Pollen and Chromatin Mutants.

**Supplemental Figure 5.** Chromatin Regulation of Pollen-Specific Gene Transcription.

**Supplemental Figure 6.** Defining the Minimum Length of Nonrandom Poly(A) Tail for the *Arabidopsis* Genome.

**Supplemental Data Set 1.** List of Retrogenes and Parents Identified by Genome-Wide Searches in *Arabidopsis*.

**Supplemental Data Set 2.** Putative *Arabidopsis* Retrogenes and Parents Not Considered in This Study.

**Supplemental Data Set 3.** Repeated Retroposition Events in *Arabidopsis*.

**Supplemental Data Set 4.** Retrogene and Parent Transcription in 49 *Arabidopsis* Developmental Stages and Tissues.

**Supplemental Data Set 5.** Transcription of Specific Gene Sets in the Wild Type and Chromatin Mutants.

**Supplemental Data Set 6.** Transcription Analysis of DNA Duplicated Genes (Blanc and Wolfe, 2004).

**Supplemental Data Set 7.** List of Genome-Wide Head-to-Head Oriented Genes and Transposable Elements.

**Supplemental Data Set 8.** Transcription of Retrogene and Head-to-Head Oriented Neighboring Genes in 49 *Arabidopsis* Developmental Stages and Tissues.

**Supplemental Data Set 9.** Transcription Quantiles of Specific Sets of Genes.

**Supplemental Data Set 10.** The Frequency of Chromatin Modifications at Specific Groups of Genes.

**Supplemental Data Set 11.** Genes Significantly Upregulated and Downregulated in *Arabidopsis* pMDC7-DUO1 Lines after 6, 12, and 24 h of DUO1 Induction.

## REFERENCES

**Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

**Arabidopsis Genome Initiative** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

**Bai, Y., Casola, C., and Betrán, E.** (2008). Evolutionary origin of regulatory regions of retrogenes in Drosophila. BMC Genomics **9:** 241.

**Baubec, T., Dinh, H.Q., Pecinka, A., Rakic, B., Rozhon, W., Wohlrab, B., von Haeseler, A., and Mittelsten Scheid, O.** (2010). Cooperation of multiple chromatin modifications can generate unanticipated stability of epigenetic states in *Arabidopsis*. Plant Cell **22:** 34–47.

**Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B.** (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res. **29:** 4319–4333.

**Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V.** (2010). LINE-1 retrotransposition activity in human genomes. Cell **141:** 1159–1170.

**Blanc, G., and Wolfe, K.H.** (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16:** 1679–1691.

**Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M., and Twell, D.** (2011). The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. Plant Cell **23:** 534–549.

**Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y., and Nurminsky, D.I.** (2002). Large clusters of co-expressed genes in the Drosophila genome. Nature **420:** 666–669.

**Bouyer, D., Roudier, F., Heese, M., Andersen, E.D., Gey, D., Nowack, M.K., Goodrich, J., Renou, J.-P., Grini, P.E., Colot, V., and Schnittger, A.** (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. PLoS Genet. **7:** e1002014.

**Coleman-Derr, D., and Zilberman, D.** (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. PLoS Genet. **8:** e1002988.

**Comai, L.** (2005). The advantages and disadvantages of being polyploid. Nat. Rev. Genet. **6:** 836–846.

**Dehal, P., and Boore, J.L.** (2005). Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. **3:** e314.

De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C.E., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc. Natl. Acad. Sci. USA 110: 2898–2903.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Farrona, S., Thorpe, F.L., Engelhorn, J., Adrian, J., Dong, X., Sarid-Krebs, L., Goodrich, J., and Turck, F. (2011). Tissue-specific expression of FLOWERING LOCUS T in Arabidopsis is maintained independently of polycomb group protein repression. Plant Cell 23: 3204–3214.

Fawcett, J.A., Maere, S., and Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc. Natl. Acad. Sci. USA 106: 5737–5742.

Flagel, L.E., and Wendel, J.F. (2009). Gene duplication and evolutionary novelty in plants. New Phytol. 183: 557–564.

Gregory, R.T., and Mable, B.K. (2005). Polyploidy in animals. In The Evolution of the Genome, R.T. Gregory, ed (New York: Elsevier), pp. 427–483.

Hoffmann, R.D., and Palmgren, M.G. (2013). Epigenetic repression of male gametophyte-specific genes in the Arabidopsis sporophyte. Mol. Plant 6: 1176–1186.

Honys, D., and Twell, D. (2003). Comparative analysis of the Arabidopsis pollen transcriptome. Plant Physiol. 132: 640–652.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol. 5: R85.

Ibarra, C.A., et al. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337: 1360–1364.

Inagaki, S., Miura-Kamio, A., Nakamura, Y., Lu, F., Cui, X., Cao, X., Kimura, H., Saze, H., and Kakutani, T. (2010). Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. EMBO J. 29: 3496–3506.

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: Classifying and distinguishing between models. Nat. Rev. Genet. 11: 97–108.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res. 20: 1313–1326.

Kaessmann, H., Vinckenbosch, N., and Long, M. (2009). RNA-based gene duplication: Mechanistic and evolutionary insights. Nat. Rev. Genet. 10: 19–31.

Lafos, M., Kroll, P., Hohenstatt, M.L., Thorpe, F.L., Clarenz, O., and Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. PLoS Genet. 7: e1002040.

Larkin, M.A., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell 128: 707–719.

Loraine, A.E., McCormick, S., Estrada, A., Patel, K., and Qin, P. (2013). RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. Plant Physiol. 162: 1092–1109.

Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A., and Kaessmann, H. (2005). Emergence of young human genes after a burst of retroposition in primates. PLoS Biol. 3: e357.

Monk, D., et al. (2011). Human imprinted retrogenes exhibit non-canonical imprint chromatin signatures and reside in non-imprinted host genes. Nucleic Acids Res. 39: 4577–4586.

Mooney, M., et al. (2013). Comparative RNA-Seq and microarray analysis of gene expression changes in B-cell lymphomas of Canis familiaris. PLoS ONE 8: e61088.

Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature 460: 283–286.

Narsai, R., Howell, K.A., Millar, A.H., O'Toole, N., Small, I., and Whelan, J. (2007). Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. Plant Cell 19: 3418–3436.

Ohshima, K. (2013). RNA-mediated gene duplication and retroposons: Retrogenes, LINEs, SINEs, and sequence specificity. Int. J. Evol. Biol. 2013: 424726.

Okamura, K., and Nakai, K. (2008). Retrotransposition as a source of new promoters. Mol. Biol. Evol. 25: 1231–1238.

Pei, B., et al. (2012). The GENCODE pseudogene resource. Genome Biol. 13: R51.

Pennisi, E. (2012). Genomics. ENCODE project writes eulogy for junk DNA. Science 337: 1159–1161, 1161.

Popova, O.V., Dinh, H.Q., Aufsatz, W., and Jonak, C. (2013). The RdDM pathway is required for basal heat tolerance in Arabidopsis. Mol. Plant 6: 396–410.

Potrzebowski, L., Vinckenbosch, N., Marques, A.C., Chalmel, F., Jégou, B., and Kaessmann, H. (2008). Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biol. 6: e80.

Remm, M., Storm, C.E., and Sonnhammer, E.L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol. 314: 1041–1052.

Roudier, F., et al. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J. 30: 1928–1938.

Sakai, H., Mizuno, H., Kawahara, Y., Wakimoto, H., Ikawa, H., Kawahigashi, H., Kanamori, H., Matsumoto, T., Itoh, T., and Gaut, B.S. (2011). Retrogenes in rice (Oryza sativa L. ssp. japonica) exhibit correlated expression with their source genes. Genome Biol. Evol. 3: 1357–1368.

Sakharkar, M.K., Chow, V.T.K., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. In Silico Biol. 4: 387–393.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat. Genet. 37: 501–506.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136: 461–472.

Song, W.-Y., Choi, K.S., Kim, Y., Geisler, M., Park, J., Vincenzetti, V., Schellenberg, M., Kim, S.H., Lim, Y.P., Noh, E.W., Lee, Y., and Martinoia, E. (2010). Arabidopsis PCR2 is a zinc exporter involved in both zinc extrusion and long-distance zinc transport. Plant Cell 22: 2237–2252.

Stajich, J.E., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 12: 1611–1618.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152: 352–364.

Vinckenbosch, N., Dupanloup, I., and Kaessmann, H. (2006). Evolutionary fate of retroposed gene copies in the human genome. Proc. Natl. Acad. Sci. USA 103: 3220–3225.

**Wang, W., et al**. (2006). High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell **18:** 1791–1802.

**Yoshida, S., Maruyama, S., Nozaki, H., and Shirasu, K.** (2010). Horizontal gene transfer by the parasitic plant *Striga hermonthica*. Science **328:** 1128.

**Zhang, J.** (2003). Evolution by gene duplication: An update. Trends Ecol. Evol. **18:** 292–298.

**Zhang, Y., Wu, Y., Liu, Y., and Han, B.** (2005). Computational identification of 69 retroposons in Arabidopsis. Plant Physiol. **138:** 935–948.

**Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X.** (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS ONE **9:** e78644.

**Zhu, X., Ma, H., and Chen, Z.** (2011). Phylogenetics and evolution of Su(var)3-9 SET genes in land plants: Rapid diversification in structure and function. BMC Evol. Biol. **11:** 63.

**Zhu, Z., Zhang, Y., and Long, M.** (2009). Extensive structural renovation of retrogenes in the evolution of the *Populus* genome. Plant Physiol. **151:** 1943–1951.
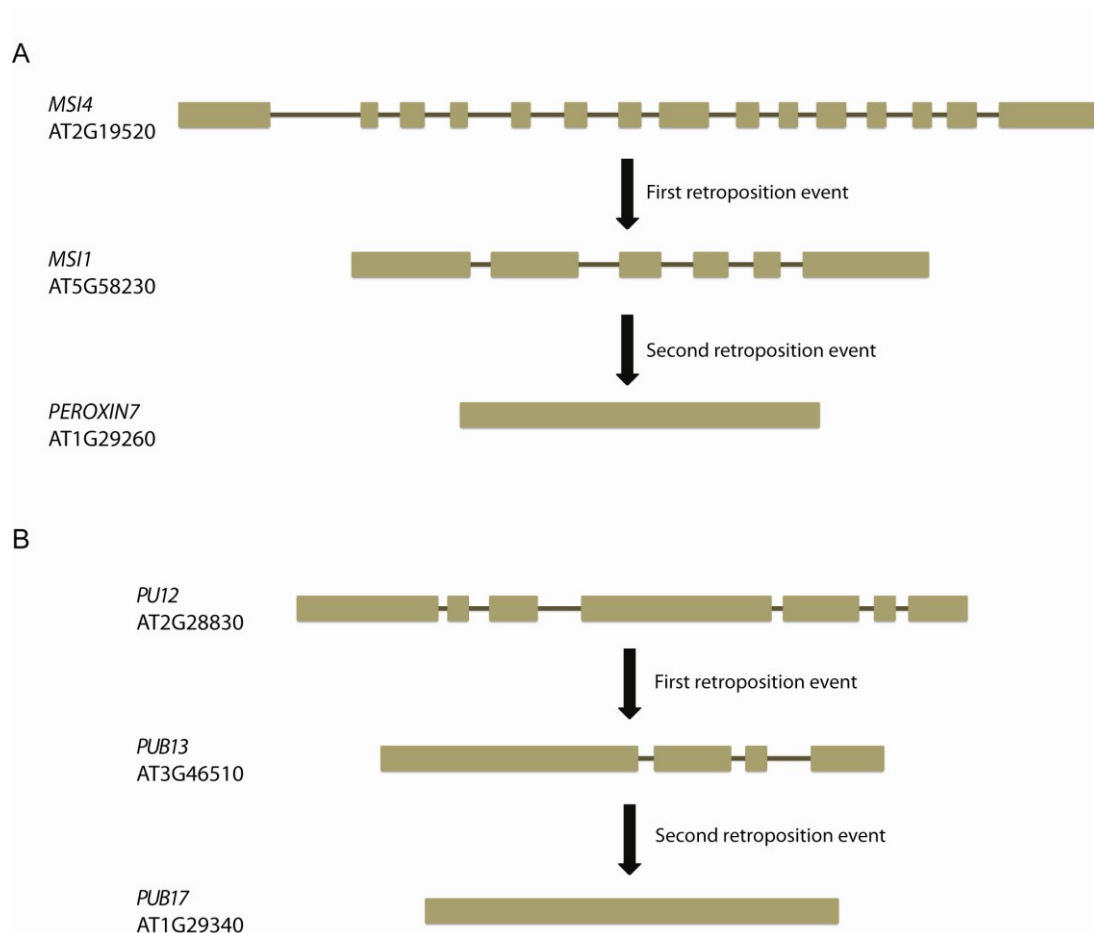
**Supplemental Figure 1. Examples of repeated retroposition in *Arabidopsis*.**
**(A)** The *MSI4 – MSI1 – PEROXIN 7* retroposition series.
**(B)** The *PUB* gene family retroposition series.
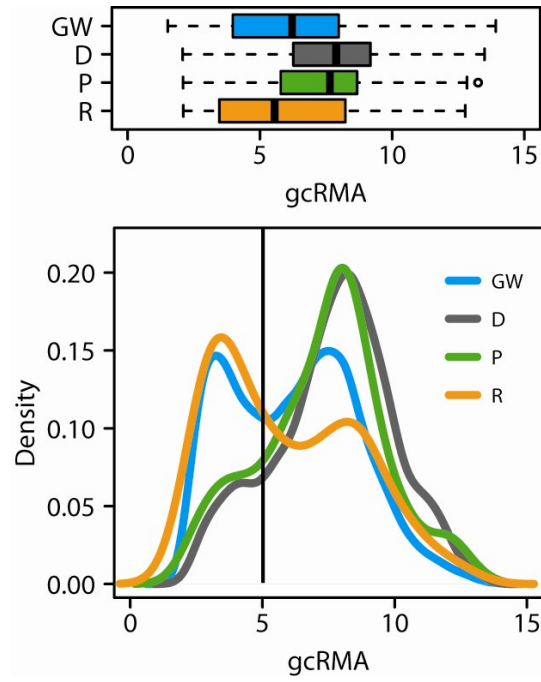
**Supplemental Figure 2. Defining lowly and highly transcribed genes in Arabidopsis.**
Box and density distribution of $\log_2$ robust multi-array averaging (gcRMA) values from 49 Arabidopsis developmental stages for all genes (GW, green), DNA duplicated genes (D, grey), parents (P, blue) and retrogenes (R, orange). The valley between low and high transcribed genes at gcRMA 5 is indicated by the vertical line.
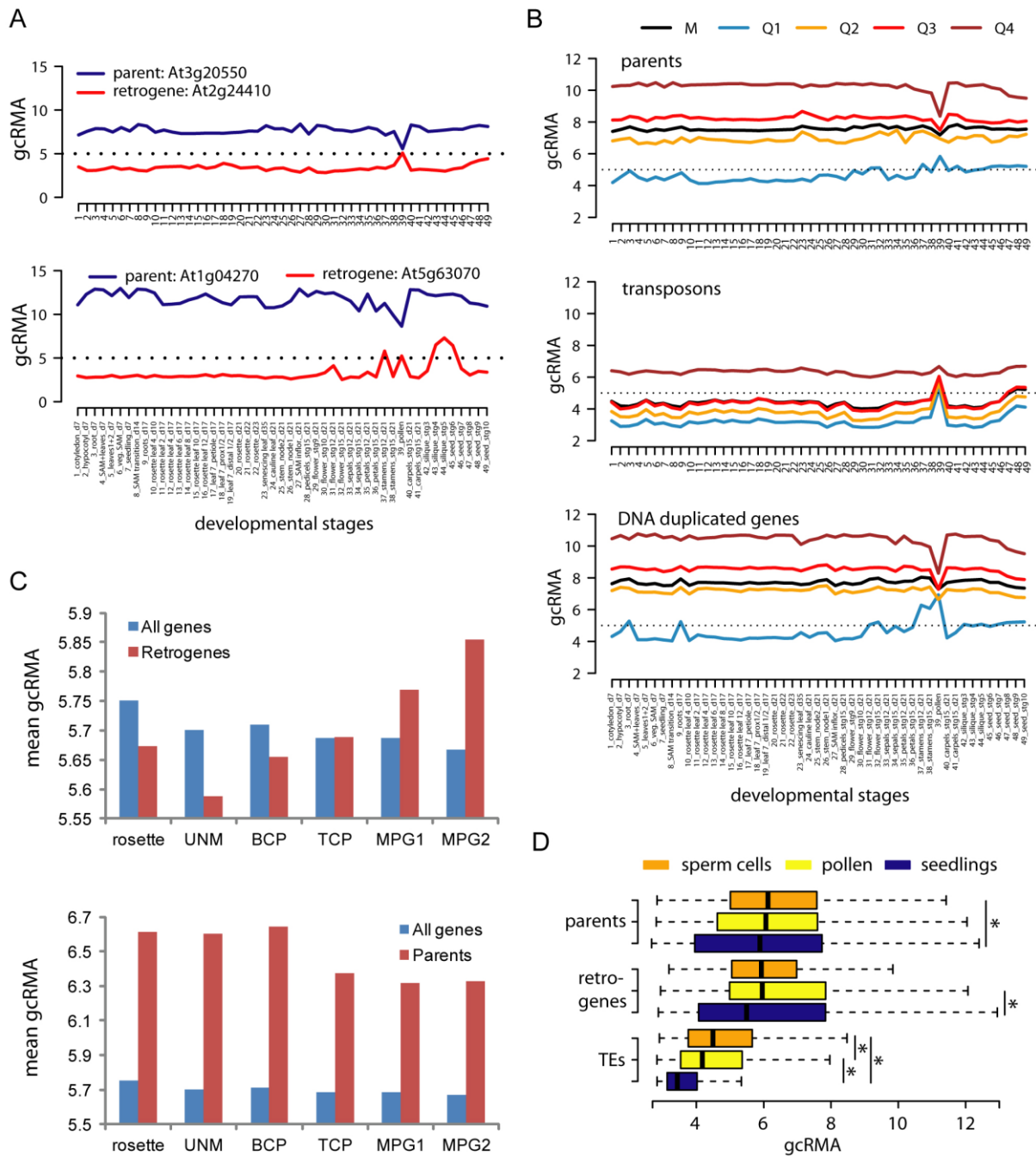
**Supplemental Figure 3. Arabidopsis retrogenes are transcribed in pollen.**

**(A,B)** Log$_2$ robust microarray averaging (gcRMA) values (*y*-axis) of specific groups of genes in 49 Arabidopsis developmental stages and tissues (*x*-axis). The horizontal dashed line (gcRMA = 5) indicates the threshold of high transcription. (A) Representative examples of retrogene-parent pairs with ubiquitously transcribed parents and tissue-specifically transcribed retrogenes. (B). gcRMA values for parents (top), transposons (middle) and DNA duplicated genes (bottom) shown as the mean (M) and transcription quantiles from low-transcribed (Q1) to high transcribed (Q4).

**(C)** Mean gcRMA transcription values for all genes, parents and retrogenes in 21 days old rosettes (rosette) and pollen developmental series consisting of: unicellular microspores (UCM), bi-cellular pollen (BCP), tri-cellular pollen (TCP) and two datasets of mature pollen grains (MPG1 and MPG1).

**(D)** gcRMA transcription values for parents, retrogenes and transposons (TEs) in pollen sperm, entire pollen (sperm cells and vegetative cells) and seedlings. Asterisks show significant differences (*P* < 0.05) in Mann-Whitney-Wilcoxon test.
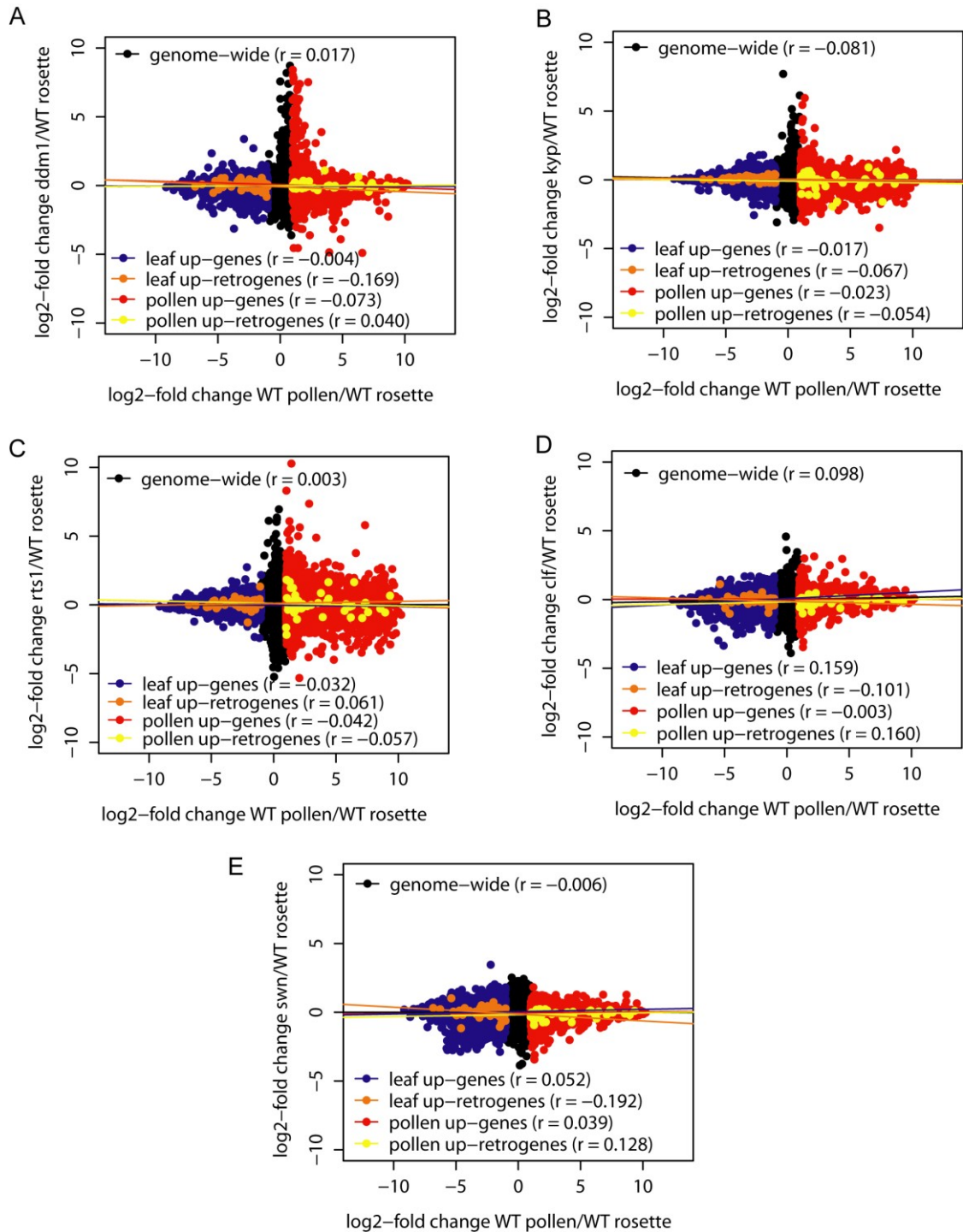
3

**Supplemental Figure 4. Transcription correlations between pollen and chromatin mutants.**

**(A-E)** Dot plots of microarray based log$_2$-fold-changes in wild type pollen (ATGE_73)/rosettes (ATGE_22) (*x*-axis) versus mutant rosettes/wild type rosettes (*y*-axis). Specific gene sets were superimposed on the genome-wide gene set. The lines indicate transcription correlation (*r*) between the *x*- and the *y*-axis gene sets. **(A)** shows comparison of pollen with *ddm1*, **(B)** with *kyp*, **(C)** with *HDA6* mutant allele *rts1-1*, **(D)** with *clf* and **(E)** with *swn*.
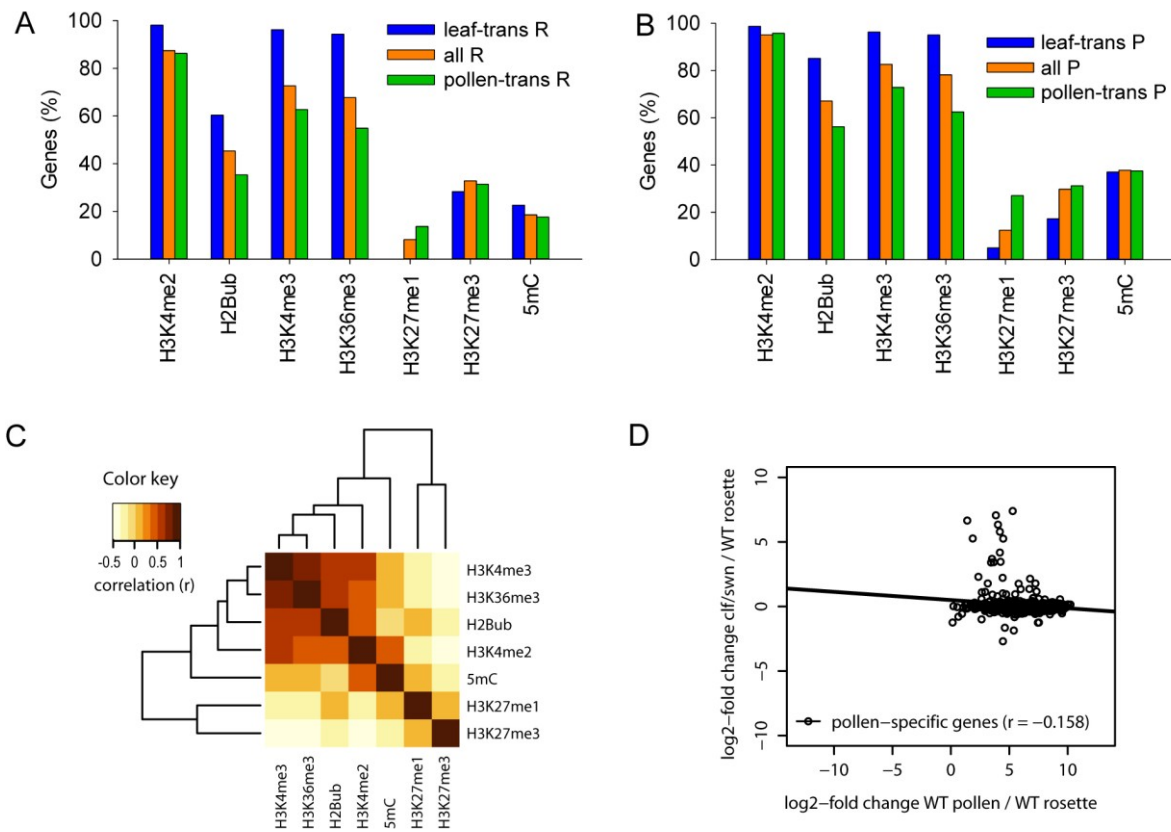
**Supplemental Figure 5. Chromatin regulationof pollen-specific gene transcription.**
**(A)** The frequency of seven chromatin modifications at protein coding regions for all retrogenes (all R), leaf-transcribed retrogenes (leaf-trans R) and pollen-transcribed retrogenes (pollen-trans R) in young leaf tissues.
**(B)** Shows the same as (A) but for parents.
**(C)** Hierarchical clustering and heat map of Pearson correlations between seven analyzed chromatin modifications for all retrogenes.
**(D)** Dot plot of $\log_2$-fold-changes in wild type pollen (ATGE73)/rosettes (ATGE_22) (*x*-axis) versus *clf/swn* doublemutant/wild type rosettes (*y*-axis) for the set of 584 pollen-specific genes defined by Hoffmann and Palmgren, 2013. The black line indicates transcription correlation (*r*) between the *x*- and the *y*-axis datasets.
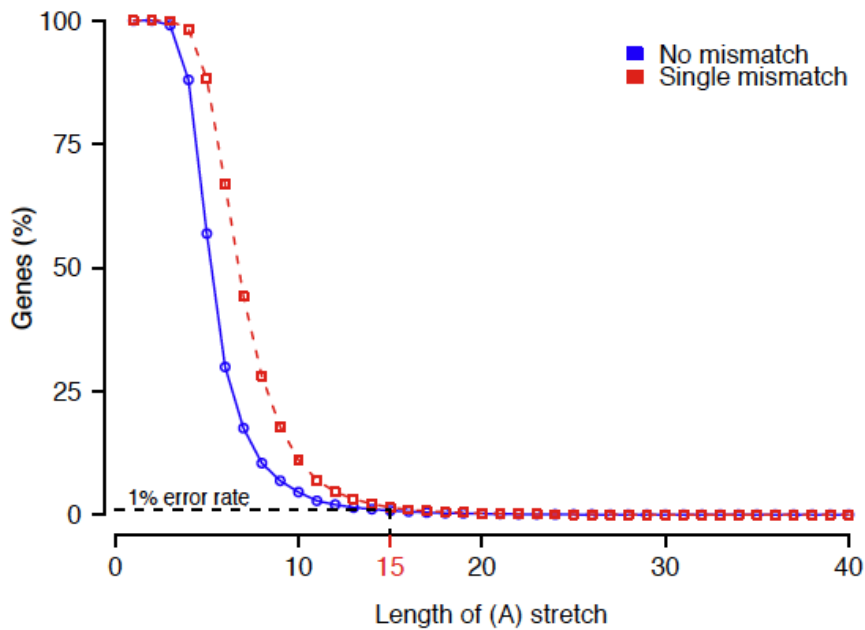
**Supplemental Figure 6. Defining the minimum length of non-random poly(A)-tail for the Arabidopsis genome.**

The length of adenine stretches (x-axis) in 150 or 250 bp downstream regions of genes with or without 3'-UTR, respectively (y-axis). Multiple adenine-stretches per gene were calculated. The 1% error rate and a single mismatch were accepted. About 99% of TAIR10 genes has stretches of adenines with a length ≤ 15 in their downstream regions. Therefore, only the genes with poly(A)-tail >15 bp were accepted as retrogene candidates.

**Publikace 9**

# Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis

Tuncay Baubec[1,2], Andreas Finke[3], Ortrun Mittelsten Scheid[1] & Ales Pecinka[1,3,*]

## Abstract

In plants, transposable elements (TEs) are kept inactive by transcriptional gene silencing (TGS). TGS is established and perpetuated by RNA-directed DNA methylation (RdDM) and maintenance methylation pathways, respectively. Here, we describe a novel RdDM function specific for shoot apical meristems that reinforces silencing of TEs during early vegetative growth. In meristems, RdDM counteracts drug-induced interference with TGS maintenance and consequently prevents TE activation. Simultaneous disturbance of both TGS pathways leads to transcriptionally active states of repetitive sequences that are inherited by somatic tissues and partially by the progeny. This apical meristem-specific mechanism is mediated by increased levels of TGS factors and provides a checkpoint for correct epigenetic inheritance during the transition from vegetative to reproductive phase and to the next generation.

## Introduction

Genomes of higher plants contain a high proportion of transposable elements (TEs). Nearly all TE families are represented with some potentially mobile copies. This endangers genome stability, especially if transposition were to occur in cells forming the germline and offspring. Plant evolution has brought about an efficient protection mechanism against extensive TE activity by preventing their expression via transcriptional gene silencing (TGS). This is mediated by epigenetic regulation through DNA methylation and repressive histone modifications [1,2]. In short, TE transcription triggers an RNA-directed DNA methylation (RdDM) mechanism that involves

de novo DNA methylation [3,4]. Depending on the sequence context, different pathways ensure correct maintenance and transmission of established DNA methylation patterns [5,6]. Conversely, disturbance of RdDM and maintenance pathways allows the transcription of specific TEs [1,7,8].

Unlike in mammals, where DNA methylation is largely erased and then re-established during germ cell maturation and zygote formation, plant DNA methylation is considered to be generally stable [1]. However, reinforced silencing of TEs has been proposed in gametes and the early embryo via mobile siRNAs produced in companion cells [9–11]. This suggests an important role for RdDM in surveying the genome of gametes and early zygotes.

Gamete formation in plants occurs late during development, and cells undergo numerous cell divisions before flowering. In addition, the formation of secondary meristems widens the range of cells that can contribute to progeny. Therefore, any loss of TE silencing during the vegetative phase [12,13] can lead to the transmission of active TEs to the next generation.

Here, we show that release of TGS control upon treatment with the DNA methylation inhibitor zebularine [14] is observed only in tissues inherited from the embryo, but not in newly developing parts of the plant. Functional analysis identifies RdDM as an important regulator of TGS maintenance in newly formed tissues, and lack thereof leads to an increased inheritance of active states to the next generation. Various meristematic tissues display enhanced expression of genes required for TGS, and we propose that this tissue-specific coordinated expression is required to enforce epigenomic stability and germline protection during vegetative growth.

## Results and Discussion

### DNA methylation inhibitors cause tissue-specific transcriptional reactivation of repetitive DNA

The cytidine analog zebularine induces transient DNA hypomethylation and transcriptional activation of otherwise silent sequences in

1  Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna, Austria
2  Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
3  Max Planck Institute for Plant Breeding Research, Cologne, Germany
   *Corresponding author. Tel: +49 221 5062 465; Fax: +49 221 5062 413; E-mail: pecinka@mpipz.mpg.de

wild-type (WT) Arabidopsis [15]. To analyze the mechanism governing re-methylation and re-silencing, we applied zebularine to the line *TS-GUS* (6b5, L5) [16] containing a transcriptionally silent β-glucuronidase transgene that is activated throughout the entire plant in the background of epigenetic mutants like *ddm1* (Fig 1A) [5]. While mock-treated plants showed no GUS staining (Fig 1B), growth in the presence of 20 or 40 μM zebularine or 400 μM 5-azadeoxycytidine released GUS silencing in cotyledons, but neither in true leaves of all stages nor in floral tissues (Fig 1C and D and Supplementary Fig S1). In addition, no GUS signal was detected in selfed progenies from zebularine-treated WT plants (Supplementary Fig S1), suggesting that the loss of silencing was restricted to embryonic tissues only. This was confirmed after zebularine application to a *TS-GFP* reporter line containing a repetitive silent GFP marker [17] that showed an even sharper separation between GFP-positive cotyledons, hypocotyl, and root and the GFP-negative true leaves around the SAM (Fig 1E and F, arrowhead).
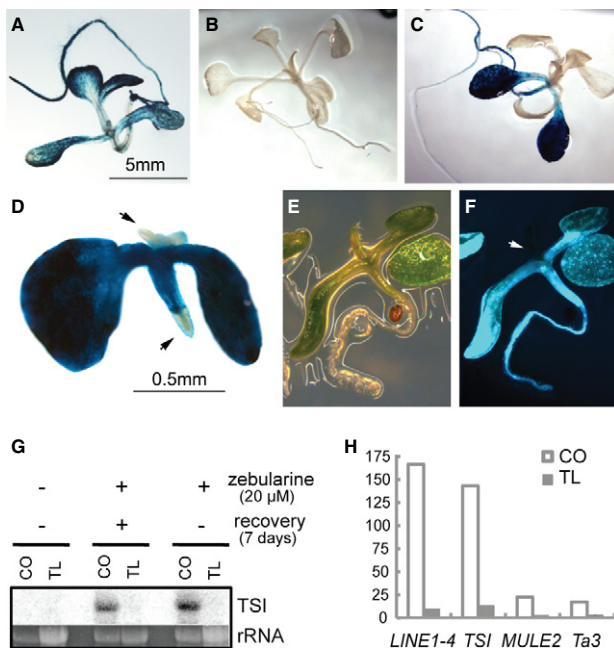


**Figure 1.  DNA methylation inhibitors induce tissue-specific reactivation of transcriptionally silenced repeats.**

A      *TS-GUS* expression in *ddm1*.

B      GUS signal is absent in mock-treated WT *TS-GUS* plants.

C, D   Tissue-specific reactivation of *TS-GUS* after treatment with (C) 40 μM zebularine for 3 weeks and (D) 400 μM 5-azacytidine for 3 days and subsequent recovery for 1 week.

E, F   Tissue-specific reactivation of a transcriptionally silent *TS-GFP* transgene after the treatment with 40 μM zebularine under visible (E) and GFP fluorescence (F) light.

G      Northern blot detection of transcription from *TSI* repeats in cotyledons (CO) and true leaves (TL) of 14-day-old mock- or 20 μM zebularine-treated seedlings.

H      Quantitative reverse transcription PCR measurements of zebularine-induced transcription of TEs in cotyledons and the first pair of true leaves of WT plants. Based on the pool of approximately 20 plants in one biological replicate.

Data information: In (D) and (F): arrows point to meristem tissues lacking *GUS* and *GFP* signals.

Zebularine-induced tissue-specific reactivation holds true also for endogenous repeats. We dissected cotyledons and the first pair of true leaves from plantlets grown for 14 days either on drug-free medium, continuously on zebularine, or on 20 μM zebularine for 7 days followed by 7-day recovery on drug-free medium. Northern blot and/or quantitative reverse transcription PCR (qRT–PCR) analysis of *Transcriptionally Silent Information* (*TSI*) repeats, *LINE1-4, MULE2,* and *Ta3* revealed no signal in mock-treated plants, while zebularine treatment—independent of recovery—released silencing only in cotyledons (Fig 1G and H).

To exclude that the lack of reactivation in true leaves was due to reduced uptake of zebularine, loss of inhibitor activity, or its dilution via DNA replication, we compared DNA methylation of centromeric repeats between cotyledons and the first pair of true leaves, for mock- and zebularine-treated plants. Methylation-sensitive Southern blots indicate that zebularine treatment reduces methylation in both tissues (Supplementary Fig S1C). Furthermore, fluorescence *in situ* hybridization (FISH) with centromeric repeat sequences revealed reduced heterochromatin condensation in nuclei from cotyledons and true leaves of inhibitor-treated plants (Supplementary Fig S1D). However, the degree of decondensation was less complete in true leaves (i.e., nuclei with full decondensation of all chromocenters), which may indicate slight differences in zebularine activity or stability in specific tissues.

Taken together, the tissue-specific activation of silent repeats after zebularine treatment argues for a regulatory mechanism that corrects the loss of TGS during early vegetative growth.

## RdDM components secure repeat silencing in true leaves in spite of inhibitor treatment

To investigate the molecular basis of the tissue-specific difference in TGS, we introgressed the *TS-GUS* transgene into mutants associated with TGS and chromatin regulation. In agreement with a previous report [5], during mock treatment, we observed full *TS-GUS* reactivation only in *ddm1* or *met1*, and none or weak cotyledon-specific activation in *cmt3, kyp, lhp1, fas1, fas2, hda6,* and RdDM mutants (Fig 2A, upper panel and Supplementary Fig S2). These tissue-specific activation patterns resembled those after zebularine treatment and prompted us to expose the low activating mutants to 20 μM zebularine, scoring for potential combinatorial effects. Remarkably, zebularine treatment led to strong GUS expression in true leaves of *ago4, drm1/drm2, drd1,* and *rdr2,* while no true leaf GUS staining was observed in *cmt3, fas1, fas2, kyp, lhp1, or hda6,* suggesting that RdDM components are involved in mediating re-silencing (Fig 2A and Supplementary Fig S2B, bottom panel).

In addition, we observed elevated true leaf-specific transcription for *LINE1-4, MULE2, and TSI* in zebularine-treated *drm1/drm2, drd1,* and *dcl3,* in comparison with treated WT plants (Fig 2B and Supplementary Fig S2). Activation of *Ta3,* an element regulated mainly by methylation of histone H3 lysine 9 and CHG [18], was not further induced in true leaves of zebularine-treated mutant plants (Fig 2B and Supplementary Fig S2). A similar response of *LINE1-4* in WT accessions Col-0 and Ws-2 indicated that this was independent of the different genetic background of the mutants (Supplementary Fig S2D).

In order to measure the combined effect of zebularine and defective RdDM on DNA methylation in true leaves, we performed
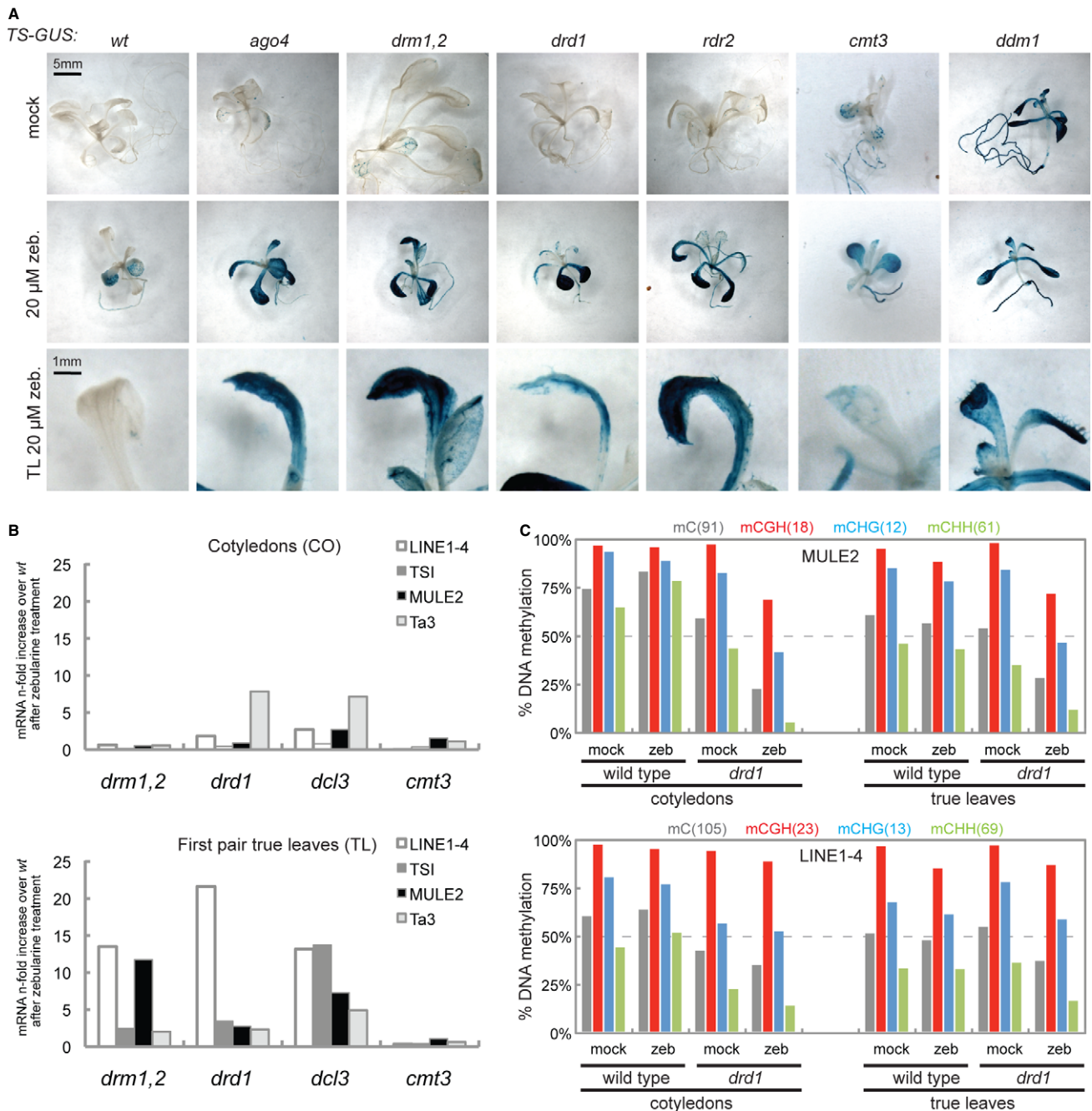
**Figure 2. Release of transcriptional silencing in true leaves upon the inhibition of DNA methylation in plants impaired in RdDM.**

A   Representative examples of GUS staining of whole seedlings (top and middle row) or their first true leaves (bottom row) after mock or zebularine treatments.

B   Quantitative reverse transcription PCR measurements of zebularine-induced reactivation from TEs in cotyledons (CO) and true leaves (TL) in transcriptional gene silencing mutants compared to WT. Based on the pool of approximately 20 plants in one biological replicate.

C   DNA methylation analysis by bisulfite sequencing. Shown is percent cytosine methylation in all sequence contexts for *MULE2* and *LINE1-4* in true leaves and cotyledons of mock- and zebularine-treated WT and *drd1*. A minimum of 15 unique clones were scored per experiment.

Source data are available online for this figure.

bisulfite sequencing at defined copies of *LINE1-4* and *MULE2* in WT and *drd1* plants (Fig 2C). Compared to untreated WT plants, lack of DRD1 resulted in a reduction in methylated cytosines by 15% in cotyledons (mostly at CHG and CHH), while in true leaves, only minor changes could be measured (< 10% reduction). Importantly, additional zebularine treatment in *drd1* plants resulted in a more pronounced hypomethylation compared to zebularine treatment or lack of *drd1* alone (Fig 2C). The additive effect of zebularine—

although at low dose—exceeded the 15% reduction at both analyzed targets in true leaves, compared to untreated WT controls, and affected all sequence contexts (Fig 2C). Surprisingly, we observed that asymmetric methylation increased after zebularine treatment in cotyledons (Fig 2C).

To further validate that RdDM antagonizes DNA methylation interference by zebularine, we germinated WT and *drd1 TS-GUS* plants on zebularine-free medium and transferred the seedlings after 6 days to zebularine-containing medium for additional 6 days (Fig 3A). Owing to the dependence on DNA replication, zebularine-mediated reactivation was observed only in tissues that proliferated during drug treatment such as newly grown parts of the root (Fig 3A). Thus, absence of replication in developed hypocotyl, cotyledons, and adult root regions protected against zebularine-mediated reactivation (Fig 3A). In addition, the lack of *drd1* resulted in TS-GUS reactivation in true leaves, validating that RdDM antagonizes the effect of zebularine treatment in true leaves (Fig 3A).
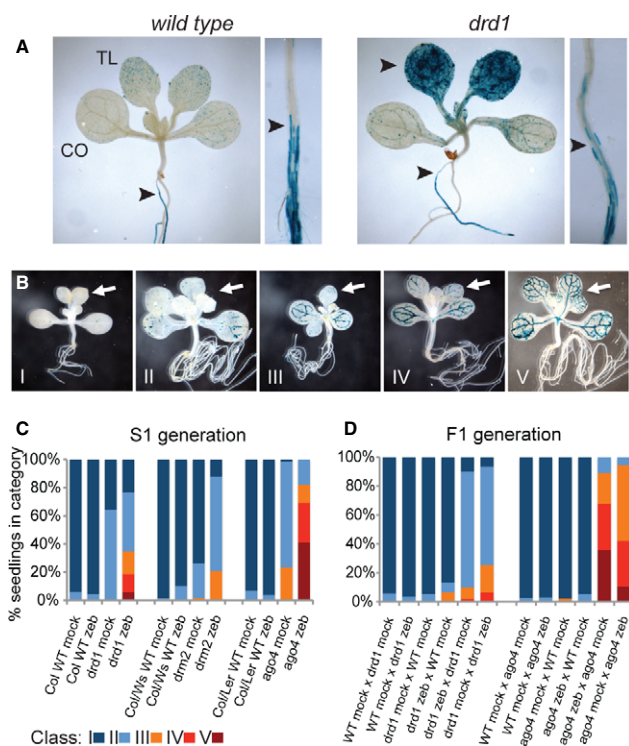


**Figure 3.   Zebularine treatment of RdDM mutants results in stochastic and transgenerational inheritance of active repetitive elements.**

A   *TS-GUS* reactivation in WT and *drd1* after 6-day mock treatment followed by 6-day 20 μM zebularine treatment. Arrowheads indicate zebularine-mediated reactivation in tissues grown in the presence of the inhibitor.

B   Representative examples of selfed (S1) seedlings from zebularine-treated S0 mutant plants with varying degree of *TS-GUS* reactivation in cotyledons and true leaves (I–V). Categories show plants with increasing degree of *TS-GUS* reactivation in true leaves (arrows).

C, D   Percentage of S1 seedlings with varying GUS reactivation from S0 mock-treated and 20 μM zebularine-treated plants (C). Percentage of F1 seedlings with varying GUS reactivation from crosses between WT and mutant plants with or without zebularine (D). Approximately 100 seedlings were scored per genotype and treatment. The classification is based on (B).

Taken together, these results indicate that release of repression from a subset of TEs in true leaves requires simultaneous chemical interference with DNA methylation and genetic interference with the RdDM pathway. Hence, we suggest a central role of RdDM in mediating re-silencing of TEs in vegetative tissues by correcting for induced inefficiency in TGS maintenance. In contrast to TE re-silencing after genetic deletions of *DDM1* that usually requires several generations [19], partial removal of methylation by zebularine is restored immediately.

**Lack of RdDM components allows the inheritance of inhibitor-activated states**

To address the inheritance of reactivated states, we screened *TS-GUS* activity in adult tissues that developed after recovery from zebularine treatment. Except for noticeable *TS-GUS* activity in the rosette leaves of 3-week-old *drd1* and in the vascular system of 5-week-old *ago4* plants, GUS expression in the remaining mutant lines was restricted to rare sectors varying in shape, size, and position between individual plants (Supplementary Fig S3B). These apparently stochastic effects were also evident from quantitative mRNA measurements for the expression of endogenous TEs, where independent biological replicates showed drastic differences between mutants or between targets (Supplementary Fig S3C).

The reactivation in adult tissues made us ask whether such stochastic activation can be transmitted to the next generation. Selfed progeny (S1) of mock- and zebularine-treated WT, *drd1*, *ago4, and drm1/drm2* (S0) were grown on zebularine-free media and compared by GUS staining (Fig 3B). WT seedlings showed no staining, irrespective of the treatment, indicating full re-establishment of *TS-GUS* silencing. S1 plantlets obtained from mock-treated mutants displayed low GUS levels in cotyledons, as observed previously (Fig 2A). However, zebularine treatment of the parental plants during the first 3 weeks of vegetative growth led to an enhanced GUS staining in cotyledons of *drm1/drm2* and furthermore in true leaves of *drd1* and *ago4* S1 progeny (Fig 3B and C). The differential degree of inheritance between individual RdDM mutants stems most likely from variable strength of silencing in the parental plants. Importantly, inheritance of active GUS was found in reciprocal crosses with mock-treated *drd1* and *ago4* plants, but was abolished in crosses with WT plants (Fig 3D).

The compromised re-silencing of repeats in zebularine-treated RdDM mutants provides evidence for a safeguarding function of the RdDM pathway during vegetative growth and consequently for the next generation. Genetic deletions of *AGO4, DRD1, or DRM2* allowed the formation of clonal patches of active transgenes in later developing parts of the plant and increased the frequency of transmission of the active state to progeny. The mosaic-like expression patterns in the progeny likely reflect incomplete demethylation in different cells during zebularine treatment, resulting in epigenetic chimeras and differential representation of the affected cells after subsequent cell divisions. Crosses with WT plants providing functional AGO4 and DRD1 could prevent the transmission of the activated state to the next generation, demonstrating the requirement of the RdDM pathway for restoring silencing at re-activated repetitive elements.

## Expression of RdDM and chromatin regulator genes is significantly increased in the SAM

The above observations suggested a qualitative and quantitative difference in the degree of TGS control and its reinforcement in different tissues. Based on the clear exemption of the meristematic region from *TS-GFP* and *TS-GUS* activation in zebularine-treated seedlings (Fig 1D and F), we argued that the SAM could play a primary role in mediating this tissue-specific response. We compared gene expression in the vegetative SAM, cotyledons, true leaves, and a set of 49 different tissues in published ATH1 microarray data [20]. First, we focused on a set of 16 genes known to be involved in TGS (Fig 4A). All of them had highest expression levels in the SAM

sample, compared to cotyledons, true leaves, or average intensities calculated across all tissues (Fig 4A, and validated by qRT–PCR for a subset of genes; Supplementary Fig S4). This indicated that the stringent silencing observed in true leaves might originate from a high abundance of TGS factors in the SAM. This is in agreement with gene expression analysis of cells in the *Arabidopsis* shoot apical stem cell niche [21]. In contrast, a control group of housekeeping genes failed to show similar differences between the analyzed samples (Supplementary Fig S4B).

We observed that genes involved in the maintenance of TGS (*MET1, DDM1, CMT3,* or *FAS1*) were less expressed in cotyledons compared to true leaves, most likely owing to lower proliferation rates in cotyledons. Nevertheless, a direct comparison between both
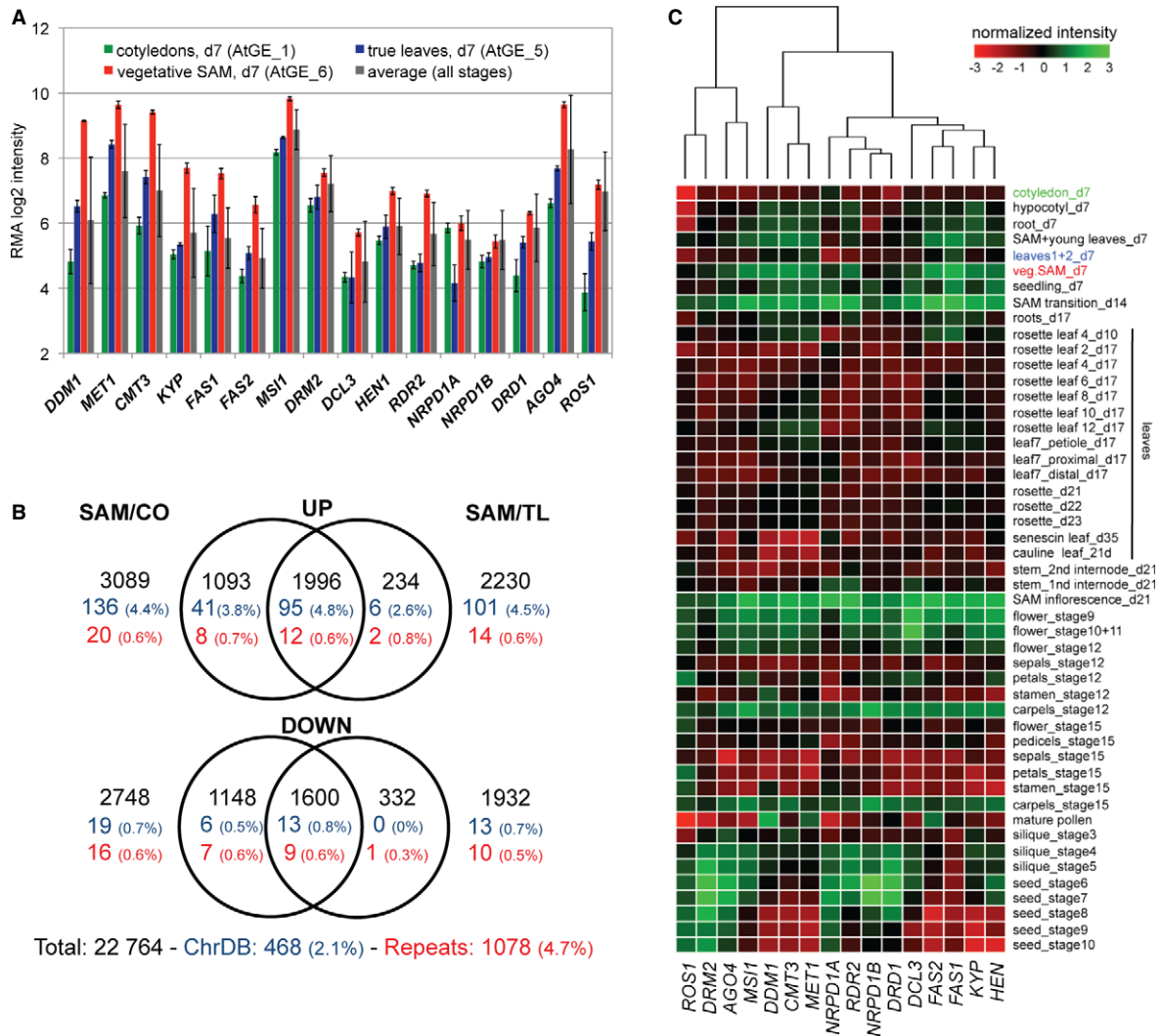


**Figure 4.  Genes connected with transcriptional gene silencing show higher expression in meristematic tissues.**

A   Robust multiarray averaging (RMA)-normalized expression array values from probe sets corresponding to selected chromatin regulator genes in 7-day-old cotyledons (green), true leaves (blue), and vegetative shoot apical meristem (SAM) (red) compared to the average intensity in 49 different tissues or developmental stages (gray, see also C). Standard deviation from three biological replicates (individual tissues) or across all arrays (average) is indicated.

B   Venn diagrams representing significant differential gene expression between SAM and cotyledons or true leaves. The total number of differentially expressed probes and the percent (in parentheses) are shown.

C   Heatmap and hierarchical clustering visualizing normalized tissue-specific expression of selected TGS genes across 49 different *Arabidopsis* tissues or developmental stages.

tissues failed to detect significant differences in the expression of additional genes involved in chromatin regulation between cotyledons and true leaves (Supplementary Table S1 and Fig 4C). We next explored whether the elevated expression of chromatin regulators in the SAM tissue was due to a global increase in transcription by directly comparing expression to cotyledons or true leaves. No significant difference was found in the number of up- and down-regulated protein-coding genes, suggesting that the overall transcriptional activity is not elevated in SAM tissues (Fig 4B and Supplementary Table S2). However, genes encoding chromatin regulators were sixfold enriched in the SAM up-regulated (4.4%) versus the SAM down-regulated probe sets (0.7%) (Fig 4B; Supplementary Fig S4C and E and Supplementary Table S2). Nevertheless, the lower abundance of chromatin regulators in young and adult leaf tissues is sufficient to maintain TGS under standard conditions (Fig 4B; Supplementary Fig S4D and Supplementary Table S2).

Besides the vegetative SAM, we observed coordinated and increased expression of genes involved in establishment and maintenance of TGS in other meristematic tissues with rapidly dividing cells, including all apical meristems at different developmental stages (e.g., vegetative growth and transition to flowering), early stages of flower development, and all stages of carpel development (Fig 4C and Supplementary Fig S4F). In contrast, tissues growing mostly by cell expansion, such as hypocotyl, stem internodes, cotyledons, or differentiated leaves, had generally lower expression from the same set of genes.

Taken together, this suggests that the elevated expression of the RdDM pathway and other chromatin regulators in meristems functions as a relay mechanism that ensures correct propagation of silent states to new tissues and organs, including the germline. Cell-type specific differences in TGS were previously reported for gametophytes and early embryonic phases of plant development where specific components of TGS are coordinately up- or down-regulated in terminally differentiated companion cells [9–11]. Although still a matter of debate [22], it has been proposed that this could lead to the generation of small RNAs complementary to TEs that reinforce silencing in the germline [9,11]. This suggests silencing checkpoints throughout gametogenesis and seed development. Preferential reinforcement of silencing in meristematic tissues, as reported here, would present a similar checkpoint during vegetative growth prior to formation of the next generation. Combined action of all three checkpoints could provide a robust surveillance mechanism that ensures silencing of TEs during vegetative growth and sexual propagation.

# Materials and Methods

### Plant material, growth conditions, and chemical treatments

The *A. thaliana* Col-0 *TS-GUS* (L5, 6b5) line [5,16] was crossed with the mutants: *rdr2-1* [23]; *drd1-6* [24]; *kyp* (SALK 041474); *fas1* (SAIL_662.D10); *fas2* (SALK_033228); *hda6* allele *rts1-1* [25]; *cmt3* in Ws-2 [26], *drm1/drm2* double mutant in Ws-2 [7]; *ago4-1* in L*er*1 [27] and *ddm1-5* in Zh [28]. The segregating F2 plants were genotyped and lines homozygous for the *TS-GUS* locus and the mutations or the WT alleles were used for analyses. Plant treatments were performed as described [15]. In brief, sterilized seeds were grown on

agar-solidified germination medium containing 20 or 40 μM zebularine (Sigma) in growth chambers under 16-h light/8-h dark cycles at 21°C. Recovery was allowed after transferring zebularine-treated seedlings to drug-free medium or soil. 5-Azacytidine (Sigma) treatment was performed by germinating seeds for 3 days in water containing 400 μM 5-azacytidine (refreshed every 24 h) and subsequent recovery on drug-free medium for 7 days. To analyze the long-term effects, three plants with or without zebularine treatment were transferred to soil and analyzed 3 weeks later for *TS-GUS* activation in cauline leaves. Their seeds (S1) were grown on zebularine-free media and analyzed.

### GUS and GFP detection and FISH

GUS staining was performed as described [15]. Samples were analyzed using a Leica MZ16FA binocular microscope with a Leica DFC300FX CCD camera. GFP was analyzed under UV illumination with a Leica GFP1 filter (425/60–480 nm). Nuclei were isolated, centromeric repeat probes prepared and FISH performed as described [15]. Images were acquired using a Zeiss Axioplan 2 microscope.

### Tissue dissection, nucleic acid isolation, and gel-blots

Cotyledons, the first pair of true leaves, and tissues enriched for SAM and RAM were dissected from 2-week-old seedlings grown either on 20 μM zebularine or on drug-free medium. DNA was extracted with Phytopure (GE Healthcare) and RNA with RNeasy (Qiagen). Gel blot analyses were performed as described [15].

### Bisulfite sequencing

DNA was isolated from cotyledons and true leaves of WT and *drd1* grown on mock and 20 μM zebularine-containing media for 14 days. Samples were bisulfite-treated by EpiTect kit (Qiagen), and *MULE2* and *LINE1-4* (Chr_2:6,881,271-6,881,800; Chr_2:378,248-378,792, respectively) were amplified from the converted DNA using primers listed in the Supplementary Table S4. At least 15 unique reads per sample were analyzed by CyMATE [29].

### qRT–PCR

DNase I-treated RNA was reverse-transcribed with random hexamer primers using RevertAid MuLV-RTase, RNaseH- (MBI Fermentas). qRT–PCR was done with SensiMix Plus SYBR and Fluorescein kit (Quantace) in an iQ5 system (Bio-Rad). PCR primers are given in Supplementary Table S4. Relative mRNA abundance was normalized to *EIF4A1* or *ACTIN2* mRNA.

### Microarray data analysis

Affymetrix ATH1 gcRMA-normalized data [20] were downloaded from http://www.weigelworld.org. Heatmaps for selected chromatin regulators were generated according to z-scores across all samples, allowing hierarchical clustering using the heatmap.2 package in R. Changes in gene expression were calculated by contrasting vegetative_SAM_d7 (AtGE_6) to cotyledons_d7 (AtGE_1) or true_-leaves_d7 (AtGE_5) in R using the Limma package. Only log$_2$

fold-changes > 1 or < −1 with adjusted *P* values < 0.05 were considered significant. 475 and 1155 probe sets corresponding to chromatin regulators (Chromatin Data Base, http://www.chromdb.org/) and TEs [11], respectively, were considered (Supplementary Table S3).

## Author contributions

TB, AF, and AP conceived, performed, and analyzed the experiments; AP, TB, and OMS wrote the manuscript.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

1. Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. *Science* 330: 622−627

2. Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. *Nature* 447: 418−424

3. Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, McCombie RW, Lavine K, Mittal V, May B, Kasschau KD et al (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430: 471−476

4. Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJM (2007) Targets of RNA-directed DNA methylation. *Curr Opin Plant Biol* 10: 512−519

5. Elmayan T, Proux F, Vaucheret H (2005) Arabidopsis RPA2: a genetic link among transcriptional gene silencing, DNA repair, and DNA replication. *Curr Biol* 15: 1919−1925

6. Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJM, Matzke M (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J* 25: 2828−2836

7. Cao X, Jacobsen SE (2002) Role of the Arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr Biol* 12: 1138−1144

8. Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol* 1: e67

9. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T et al (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* 337: 1360−1364

10. Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature* 460: 283−286

11. Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136: 461−472

12. Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Scheid OM (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell* 22: 3118−3129

13. Tittel–Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I (2010) Stress-induced activation of heterochromatic transcription. *PLoS Genet* 6: e1001175

14. Zhou L, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP (2002) Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J Mol Biol* 321: 591−599

15. Baubec T, Pecinka A, Rozhon W, Mittelsten Scheid O (2009) Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *Plant J* 57: 542−554

16. Morel J-B, Mourrain P, Béclin C, Vaucheret H (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr Biol* 10: 1591−1594

17. Blevins T (2009) *Molecular Genetic Analysis of siRNA Biogenesis and Function in Arabidopsis thaliana.* PhD Thesis, University of Basel, Switzerland. http://oatd.org/oatd/record?record=oai\:edoc.unibas.ch\:1029

18. Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416: 556−560

19. Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF et al (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* 323: 1600−1604

20. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37: 501−506

21. Yadav RK, Girke T, Pasala S, Xie M, Reddy GV (2009) Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. *Proc Natl Acad Sci USA* 106: 4941−4946

22. Grant–Downton R, Kourmpetli S, Hafidh S, Khatab H, Le Trionnaire G, Dickinson H, Twell D (2013) Artificial microRNAs reveal cell-specific differences in small RNA activity in pollen. *Curr Biol* 23: R599−R601

23. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* 2: e104

24. Kanno T, Mette MF, Kreil DP, Aufsatz W, Matzke M, Matzke AJM (2004) Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr Biol* 14: 801−805

25. Aufsatz W, Mette MF, van der Winden J, Matzke M, Matzke AJM (2002) HDA6 a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J* 21: 6832−6841

26. Bartee L, Malagnac F, Bender J (2001) Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes Dev* 15: 1753−1758

27. Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 Control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299: 716−719

28. Mittelsten Scheid O, Afsar K, Paszkowski J (1998) Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. *Proc Natl Acad Sci USA* 95: 632−637

29. Hetzl J, Foerster AM, Raidl G, Mittelsten Scheid O (2007) CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J* 51: 526−536

# Publikace 10

# Stress-Induced Chromatin Changes: A Critical View on Their Heritability

Ales Pecinka[1,*] and Ortrun Mittelsten Scheid[2]

[1]Max Planck Institute for Plant Breeding Research, Cologne, Germany
[2]Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna, Austria
*Corresponding author: E-mail, pecinka@mpipz.mpg.de, Fax, +49-221-5062-413.

**The investigation of stress responses has been a focus of plant research, breeding and biotechnology for a long time. Insight into stress perception, signaling and genetic determinants of resistance has recently been complemented by growing evidence for substantial stress-induced changes at the chromatin level. These affect specific sequences or occur genome-wide and are often correlated with transcriptional regulation. The majority of these changes only occur during stress exposure, and both expression and chromatin states typically revert to the pre-stress state shortly thereafter. Other changes result in the maintenance of new chromatin states and modified gene expression for a longer time after stress exposure, preparing an individual for developmental decisions or more effective defence. Beyond this, there are claims for stress-induced heritable chromatin modifications that are transmitted to progeny, thereby improving their characteristics. These effects resemble the concept of Lamarckian inheritance of acquired characters and represent a challenge to the uniqueness of DNA sequence-based inheritance. However, with the growing insight into epigenetic regulation and transmission of chromatin states, it is worth investigating these phenomena carefully. While genetic changes (mainly transposon mobility) in response to stress-induced interference with chromatin are well documented and heritable, in our view there is no unambiguous evidence for transmission of exclusively chromatin-controlled stress effects to progeny. We propose a set of criteria that should be applied to substantiate the data for stress-induced, chromatin-encoded new traits. Well-controlled stress treatments, thorough phenotyping and application of refined genome-wide epigenetic analysis tools should be helpful in moving from interesting observations towards robust evidence.**

**Keywords:** Chromatin • Evolution • Stress • Transgenerational stress memory.

**Abbreviations:** ARP6, actin-related protein 6 (subunit of SWR1); CAF-1, chromatin assembly factor 1; FAS1,2, fasciated 1,2 (subunits of CAF-1); H2A, histone H2A (canonical nucleosome subunit); H2A.Z, histone variant H2A.Z; MSAP, methylation-sensitive amplified polymorphism; qPCR, quantitative PCR; SWI2/SNF2, SWItch2/Sucrose Non-Fermentable2 (remodeling complex); SWR1, Swi2/Snf2-related 1 (remodeling complex); TEs, transposable elements; TGS, transcriptional gene silencing; TSI, transcriptionally silent information (repetitive genomic sequence)

## Introduction

Stress, in a biological context, refers to the consequences if organisms fail to respond adequately to unfavorable conditions. If stress cannot be avoided, e.g. by hiding or migration, physiological reactions are activated that help protect the organisms against deleterious effects, although a substantial impact on fitness, growth and development is often unavoidable. Plants, as sedentary organisms, have developed an impressive portfolio of stress responses. Nevertheless, pathogen attacks, drought, salinity or extreme temperatures can have a significant impact on vigor, including biomass production and yield in agriculture. Therefore, progress in plant breeding and biotechnology towards more stress-resistant cultivars requires better understanding of plant stress responses, to reduce such losses. Moreover, the need for greater insight into the stress defense mechanisms of plants will increase with the predicted rise of average temperatures and longer periods of extreme weather (Ahuja et al. 2010). The challenges of these changes will not only affect cultivated plants but will also have a tremendous impact on whole ecosystems including wild species. Thus, studying plant responses to abiotic stress may also be helpful in understanding plant ecology and evolution, the disappearance of species and colonization of new niches often with unfavorable conditions.

Approaches to understanding stress responses have been the focus of plant biologists for a long time and have provided extensive knowledge about various physiological stress responses and their molecular bases (Chinnusamy et al. 2004, Yamaguchi-Shinozaki and Shinozaki 2006, Huang et al. 2012).

The early phases, and specificity, of stress perception have been of special interest to researchers, as these determine subsequent downstream reactions. Also, the return to the pre-stress physiology, once the adverse conditions are gone, has been well investigated. However, the long-term perspective, addressing the potential for a 'stress memory' or heritability of stress effects in case of lasting effects, is less well studied. This originates from the general consensus that most traits determining stress resistance have a genetic basis and are subject to Darwinian natural selection and Mendelian inheritance. While there is no doubt about the validity of these principles, supported by the successful introgression of stress resistance traits during plant breeding, the occasional rapid development of new, sometimes unstable, traits is not easily reconciled with this concept (Jablonka and Raz 2009). Therefore, other, 'faster' mechanisms for long-term adaptation have been postulated and often related to the idea of Lamarckian inheritance, assuming that 'an organism can pass on characteristics or potential that it acquired during its lifetime to its offspring' (http://en.wikipedia.org/wiki/Inheritance_of_acquired_characteristics). For a long time, this idea was rejected for two main reasons. First, in spite of many attempts, a Lamarckian type of inheritance could not be reproducibly confirmed. Secondly, the concept was heavily misused to perform pseudo-scientific experiments and eliminate the opponents of Trofim Lysenko and his colleagues in the first half of the last century in Russia. However, in recent years, the development of highly sensitive stress reporter systems and the discovery of epigenetic mechanisms have revived the idea of Lamarckian 'fast' inheritance (Koonin and Wolf 2009). Indeed, some epigenetic phenomena, e.g. paramutation (reviewed in Chandler and Stam 2004), lead to the quick loss or gain of novel phenotypes that are inherited in a non-Mendelian manner. Yet, although the genetic and molecular basis of paramutation is quite well understood and in agreement with classical paradigms, a connection with stress response is not obvious. Perception of stress in one part of the plant can cause increased resistance throughout the whole plant in the process of systemic acquired resistance, and, in a process termed priming, slight stress exposure of plants leads to faster and better responses upon subsequent, more severe treatments. Again, both phenomena are explicable by molecular effects on stress perception and signaling components (reviewed in Shah 2009, Conrath 2011), and there is no evidence for their transmission to the next generation. A more likely carrier of heritable information is chromatin, the complex of genomic DNA with specialized proteins that determine the organization and packaging of the long DNA molecules within the nucleus. DNA is wrapped around nucleosomes, which are abundant chromatin protein octamers consisting of $2 \times 4$ different histone molecules. The N-terminal tails of the histones protrude from the spherical nucleosomes and can be covalently modified by acetylation, methylation, phosphorylation, ubiquitination and other residues. Also the cytosine residues of DNA can be methylated. All modifications together change the physical and chemical properties of genomic DNA. Chromatin controls the accessibility for DNA-interacting factors via condensation and provides information about gene expression potential in an epigenetic manner, i.e. in addition to DNA sequence information. Disturbances of chromatin structure result in de-regulation of gene transcription or hypersensitivity to DNA damage and can lead to abnormal development. As will be described below, there is growing evidence that stress responses can directly or indirectly modify epigenetic regulation and chromatin. As some chromatin changes are stable and become independent of the trigger, and in extreme cases form heritable epialleles (Cubas et al. 1999, Soppe et al. 2000, Manning et al. 2006), it is conceivable that stress induces persistent, or even heritable, chromatin modifications that alter gene expression and phenotypic traits, and thereby overrides Darwinian selection based exclusively on genome information. Here, we review recent literature on plant chromatin responses to abiotic stimuli and stress, their duration and functional significance, and discuss the criteria to claim their heritability.

## Chromatin changes in response to stress

### Short-term and transient responses

Reports on chromatin modifications upon external stimuli are numerous and diverse. Among abiotic stress factors, the best documentation exists for the effects of heat, which causes epigenetic deregulation and transposon activation (Lang-Mladek et al. 2010, Pecinka et al. 2010, Tittel-Elmer et al. 2010). This requires severe conditions and a certain duration of heat exposure, and it is enhanced by preceding cold treatment (Tittel-Elmer et al. 2010). The response is associated with loss of DNA-bound nucleosomes and transient heterochromatin de-condensation (Pecinka et al. 2010). Less drastic heat exposure affects histones more specifically: the transcript profile of mutants lacking *ACTIN RELATED PROTEIN 6* (*ARP6*) resembles that of heat-exposed plants even at ambient temperature (Kumar and Wigge 2010). ARP6 is part of the SWI2/SNF2 nucleosome assembly complex required for loading the histone H2A.Z variant onto DNA predominantly at transcriptional start sites. H2A.Z nucleosomes are more tightly associated with DNA than nucleosomes with canonical H2A but become evicted by higher temperature. Loss of ARP6 function mimics the state after heat-induced H2A.Z dissociation and thereby results in similar transcriptional regulation and phenotypes. Thus, H2A.Z-mediated regulation of gene expression incorporates a thermo-sensing signal and represents a bona fide functional chromatin response to a change of an abiotic parameter (Kumar and Wigge 2010).

Heat, but also other abiotic stress types, leads to transcriptional activation of several transgenic and endogenous targets of transcriptional gene silencing (TGS)—a mechanism controlling repression and heterochromatinization of repetitive DNA regions in plants (reviewed in Madlung and Comai 2004, Chinnusamy and Zhu 2009, Mirouze and Paszkowski 2011,

Paszkowski and Grossniklaus 2011, Khraiwesh et al. 2012). Here we focus on several recent studies with Arabidopsis, so far providing the deepest insight into *cis*- and *trans*-acting factors and mechanisms. Genome-wide expression analysis after prolonged heat or cold–heat stress revealed significant transcriptional up- and down-regulation of 1–2% of approximately 1,500 transposable elements (TEs) represented by probe sets on the ATH1 microarray (Pecinka et al. 2010, Tittel-Elmer et al. 2010). All TEs returned to their pre-stress expression level within <2 d of recovery at ambient temperature, with the exception of the *COPIA78* retrotransposon family. Transcripts of these TEs were detectable early (relative to other TGS targets) after onset of stress, and their high levels were still present up to 7 d post-stress. The potential for reintegration of new copies of this TE into the genome in the case of compromised epigenetic control (Ito et al. 2010, Ito et al. 2011) is discussed elsewhere (Mirouze and Paszkowski 2011, Paszkowski and Grossniklaus 2011). Chromatin analysis revealed transcriptional activation of these and other activated elements to be independent of DNA de-methylation and loss of histone H3 lysine 9 di-methylation (Lang-Mladek et al. 2010, Pecinka et al. 2010, Tittel-Elmer et al. 2010), two epigenetic marks reduced upon reactivation in the background of several TGS mutants. Instead, the genomic copies of the heat-induced TEs and many other genomic regions (including non-transcribed sequences) had reduced nucleosome occupancy, concomitant with the above-mentioned heterochromatin dissociation (Pecinka et al. 2010). A role for nucleosome loading, rather than specific modification marks, is further suggested by delayed re-silencing of heat stress-activated *TRANSCRIPTIONALLY SILENCED INFORMATION* (*TSI*), an *ATHILA*-related retrotransposon (Steimer et al. 2000), in mutants with reduced FASCIATA 1 and 2 proteins (FAS1 and FAS2), the two largest subunits of the CHROMATIN ASSEMBLY FACTOR 1 (CAF-1) (Pecinka et al. 2010). Thus, interference of prolonged heat stress with epigenetic gene silencing may be due to transient changes of nucleosome loading and chromatin organization rather than DNA or histone methylation.

A direct connection between the temperature-sensing H2A.Z at transcription-competent start sites of genes (Kumar and Wigge 2010) and the heat-induced loss of nucleosomes from heterochromatic repeats (Pecinka et al. 2010) is unlikely as DNA methylation typical for the latter is mutually exclusive with H2A.Z domains (Zilberman et al. 2008). However, both responses have in common that the removal of histones does not increase expression of all genes equally and therefore is not sufficient for transcriptional activation. The occurrence of multiple histone variants, modifications, chaperones and different nucleosome loading make it likely that chromatin dynamics upon stress are the result of a complex interplay between physical factors, their perception, pre-existing chromatin structure and maintenance mechanisms.

Like abiotic factors, pathogen-induced stress can also result in chromatin responses, and different features of chromatin affect the defense against pathogens. Infections, or chemicals mimicking pathogen attack, can change histone acetylation and methylation (Butterbrodt et al. 2006, Mosher et al. 2006, Jaskiewicz et al. 2011, Kim et al. 2012). Further, there is a correlation between the amount of a histone ubiquitin ligase and resistance to necrotrophic fungi (Dhawan et al. 2009), and loss of a histone methyltransferase results in enhanced susceptibility to bacterial pathogens (Palma et al. 2010). Involvement of chromatin remodeling in signaling of biotic stress is further suggested by decreased resistance to necrotrophic fungi of mutants with an impaired SWI/SNF ATPase (Walley et al. 2008). A role for histone variant placement is indicated by reduced salicylic acid-induced immunity in mutants lacking subunits of the SWR1 complex that installs histone variant H2A.Z (March-Diaz et al. 2008). Pathogens can also interfere with the hosts' chromatin in their favor (reviewed in Ma et al. 2011).

## Memory effects reset upon reproduction

While changes in gene expression and chromatin triggered by the stressful conditions described above are largely transient, i.e. reconstituted to the pre-stress situation shortly after return to favorable conditions, there are several processes that indicate a 'memory' effect, sometimes lasting for the lifetime of the affected individual. The best documented case in connection with a chromatin signature is the process of vernalization, i.e. the control of flowering time by preceding exposure to low temperatures. Vernalization causes repression of flowering-inhibiting factors and, once installed, this suppression persists even upon return to higher temperatures. In Arabidopsis, this involves the recruitment of chromatin-modifying enzymes to specific target genes and their subsequent inactivation (reviewed in Adrian et al. 2009, Kim et al. 2009). There is no evidence that a 'memory' of vernalization is inherited from cold-exposed individuals to the next generation, but rather there is a well-documented resetting by renewed up-regulation of the flowering inhibitor during early embryo development (Sheldon et al. 2008). In addition, the cold temperature is necessary for an important developmental switch and cannot be considered as a stress in the sense of unfavorable conditions. This is different from the case of memory effects mentioned earlier, such as systemic acquired resistance, immunity, priming or acclimation. Perception factors and signal cascades are certainly key components in these processes, but growing evidence indicates that they can result in chromatin and DNA methylation changes at specific genes which, in turn, render these genes differentially responsive to later stimuli (reviewed in Jarillo et al. 2009, van den Burg and Takken 2009, Luo et al. 2011, Ma et al. 2011, Santos et al. 2011, Yaish et al. 2011, Grativol et al. 2012, Zhu et al. 2012). The enhanced or decreased susceptibility to renewed stress and the corresponding chromatin changes can persist for different periods beyond the primary exposure, sometimes for a long time, but there is no undisputed evidence that they are stably inherited by subsequent generations.

## Lasting responses inherited by progeny

The last statement of the previous paragraph will not go unopposed, as there are numerous reports of experiments supposedly demonstrating stress-induced epigenetic states that are inherited by non-stressed progeny (recently, for example, Bilichak et al. 2012, Luna et al. 2012, Rasmann et al. 2012, Slaughter et al. 2012; more references reviewed in Boyko and Kovalchuk 2011). Rightly, these studies have received special attention as they propose a principally novel type of stress adaptation and revive the idea of inheritance of acquired characters. We, and others, have conducted an extensive literature review and identified several common issues that limit an unambiguous interpretation and acceptance of these studies. Based on this, we conclude that firm evidence for a role for chromatin modification in inheritance of stress-induced changes is still missing in plants. However, we agree that it is a very exciting field of research and, therefore, we propose criteria that we would like to see fulfilled during the analysis of trans-generational epigenetic memory effects. We believe that sharing these points with the research community may help to provide new, incontestable evidence for a direct and durable chromatin-encoded impact of environmental parameters on phenotype and adaptation.

(i) Stress-induced expression changes of *trans*-acting chromatin modifiers do not unconditionally lead to quantitative changes of the respective chromatin mark. Lower expression of the DNA methyltransferase responsible for replication-associated maintenance methylation can only be effective if the inducing conditions do not arrest the cell cycle at the same time (Steward et al. 2000). Therefore, the analysis should include transcript and protein levels (in ideal cases protein activity), accessibility of the substrates and implementation of the chromatin changes at the specific targets.

(ii) Transgenic reporter constructs for visualization of epigenetic effects have different expression levels, patterns and sensitivity, and need to be chosen carefully. The same reporter can be reactivated to various extents by different mutations (Elmayan et al. 2005), between strong expression in coherent cell lineages and weak, stochastic expression in individual cells. Transgenerational changes require the epigenetic change to occur in sectors or cells forming the germline, and must be significant enough to become permanent. Even genetically induced epigenetic switches can appear stable but revert after a few generations (Foerster et al. 2011), and lines containing transgenic homologous recombination substrates show occasional hyper-responsiveness and high variation even upon mock treatments. The variation between experiments can be of the same order as responses under inducing conditions within an experiment (Pecinka et al. 2009). Therefore, we suggest that data generated using transgenic constructs should be confirmed with experimentally different strategies, as with work with endogenous indicators, or independent quantification methods such as quantitative PCR (qPCR) analysis.

(iii) Stress in nature often consists of several components, and plants have adapted to cope with multiple stress types simultaneously, as reflected in the many signaling components involved in different stress responses (Huang et al. 2012). While researchers usually try to apply one defined stress type at a time, this might not always be successful, due to incomplete control over growth conditions, undetected pathogen infestations, difficult dosing of stress or unavoidable side effects in experiments. Lack of reproducibility and different results between labs and/or experiments can be reduced by very carefully establishing the stress conditions prior to the actual experiments, recording as many parameters as possible, and repeating experiments with the same stress treatment under otherwise slightly different settings. Any trans-generational stress memory that is relevant under highly variable conditions in nature should be robust enough to be reproduced this way.

(iv) DNA methylation is a well-established and important epigenetic mark in plants. However, it is not always the primary indicator of chromatin changes and depends in part on the level of small RNA molecules and other, already DNA-associated marks (Kanno and Habu 2011). DNA methylation differences can be indirect effects, or even be absent in spite of chromatin changes (Pecinka et al. 2010). Chromatin needs to be analyzed in a synoptic view on different features, including DNA methylation, small RNA quantification, specific histone modifications and DNA–histone association.

(v) Analysis of DNA methylation is very popular as an indicator of stress-related changes, as it is relatively easy to investigate by various methods. However, many of the techniques [e.g. cytosine extension assays, methylation-sensitive amplified polymorphism (MSAP) and Southern blots with methylation-sensitive restriction enzymes] limit the experiments to certain genomic regions and cannot quantify or detect heterogeneity of methylation. They can provide preliminary evidence for genome-wide or region-specific differences, but these should be substantiated with bisulfite sequencing, offering either locus-specific or genome-wide single base resolution (Gupta et al. 2010).

(vi) The role of DNA methylation can differ depending on its location within genes. In addition to functionally discrete modification of cytosines in different sequence contexts (CG, CHG or CHH), CG methylation in repetitive sequences, transposons and gene promoters is usually associated with transcriptional silencing, while methylated CG within exons and introns is prominent in the centre of moderately transcribed genes (reviewed in Saze and Kakutani 2011). Although the role of this gene body methylation is not clear, it is probably quite different from methylation at inactive parts of the genome (Saze and Kakutani 2011). This needs to be considered if stress-induced methylation changes are interpreted.

(vii) Correlation is not causality: stress-related phenotypes or susceptibilities may appear connected with epigenetic changes (typically DNA methylation; see point iv) but these can be secondary effects or independent spontaneous variations (Becker et al. 2011, Schmitz et al. 2011), without relevance. Claims for a causal relationship between defined changes (see points v and vi) and stress responsiveness should be

proven by analysis including mutants, overexpressors, inhibitors, etc.

(viii) In plants, trans-generational inheritance of induced chromatin changes is more difficult to define than in animals, due to the late separation of germline-forming cells from other somatic cells. Flowers containing the pre-meiotic cells are not more protected from stress exposure than other aerial plant parts, and differentiated somatic cells can re-differentiate into meristematic tissue and open a new germline via somatic embryogenesis (Verdeil et al. 2007). Re-establishment of a chromatin state after genetic interference may take more than one generation (Teixeira et al. 2009). Therefore, caution is required not to mistake such 'carryover' effects for proof of trans-generational inheritance. Claims for a memory effect should be documented by significant changes observed for more than two subsequent non-stressed generations, as in the case of the chromatin-based gene expression change in *Drosophila*, so far the best evidence for heritable effects after defined heat stress treatment (Seong et al. 2011). However, even here, the transcriptional activation is lost in the third non-stressed generation. Boosting the response by repeated treatments in subsequent generations makes the effect stronger, but not longer lasting (Seong et al. 2011), and can theoretically also be explained by additive carryover effects. Further, stress application restricted to the early part of the life cycle can help to reduce possible artifacts produced by affecting the progeny-forming cells while they are still contained within the exposed plant. At least in animals, a critical window for chromatin-related changes is limited to early developmental stages (Skinner 2011). A recent critical review of trans-generational epigenetic inheritance in mammals lists further arguments for transmission of diffusible molecules, rather than chromatin-based mechanisms (Daxinger and Whitelaw 2012).

(ix) Many experiments addressing non-genetic trans-generational inheritance are not performed with actual stress treatments, but rather with different inhibitors or toxic compounds (Guerrero-Bosagna and Skinner 2012). While some of these might be good at mimicking stress by interfering with certain components in the signaling pathways, they may however, have unnoticed side effects that would not occur with the physical or pathogen-induced stress, or they could miss some targets of those more systemic treatments. Therefore, stress-inducing or stress-mimicking drugs should be used with caution and include validation of the results with more genuine stress.

(x) Recent studies have shown that stress-induced chromatin effects can result in genetic changes (Ito et al. 2010, Ito et al. 2011, Matsunaga et al. 2012), or genetic changes can cause reprogramming of previously stable epigenetic states (Foerster et al. 2011). Any analysis of heritable chromatin change therefore needs to exclude simultaneous *trans*-acting genetic changes. With the exception of closely linked genetic and epigenetic changes, a proof of true breeding of the affected chromatin configuration upon outcrossing with non-affected plants could help to exclude such a connection.

(xi) Finally, any transmitted stress-induced chromatin change is relevant for a discussion about inheritance of acquired characters only if the change provides a benefit under specific conditions, i.e. affects the progeny's stress resistance, stress responsiveness or adaptability. Therefore, the progeny should be scored carefully for their performance under the same type of stress as applied to the ancestors, and for general fitness in comparison with progeny of unexposed plants.

According to these criteria, and to the best of our knowledge, no published data set unambiguously demonstrates trans-generational inheritance of an exclusively epigenetic and stable change induced by stress exposure of plants. Even severe conditions applied under laboratory conditions do not seem to be sufficient for permanent and/or complete erasure of pre-stress chromatin marks (Pecinka et al. 2010, Tittel-Elmer et al. 2010). Rather, a 'memory' function exists for maintaining existing or restoring disturbed chromatin states, as shown after genetic interference with DNA methylation (Teixeira et al. 2009), and not for remembering disruptions. Maintenance and restoration of chromatin states involves sophisticated, sometimes redundant and self-reinforcing mechanisms, for which quite a few components are known (Vaillant and Paszkowski 2007, Law and Jacobsen 2010, Kanno and Habu 2011, Meyer 2011, Saze et al. 2012). In addition, they can be determined by the DNA sequence itself, as shown by the autonomous installation of DNA methylation patterns independent from transcription, genomic location and neighboring sequences in fungi and mammalian cells (Miao et al. 2000, Lienert et al. 2011), or by partially sequence-determined nucleosome positioning (Segal et al. 2006, Chodavarapu et al. 2010).

## Chromatin responses to stress in evolutionary perspective

In spite of the maintenance mechanisms, chromatin undergoes a lot of programmed or induced changes upon developmental and exogenous triggers, as described above. It is evident that individual stress-related genes in plants are also partially regulated at the chromatin level. Chromatin effects on other genes or genome-wide changes upon stress are less plausible. They could contribute to stress response in an as yet unknown way, or open a 'window of opportunity' for potentially beneficial changes (including a putative stress memory for future times or generations), thereby having a selective benefit. Alternatively, undirected effects could be a 'sign of imperfection' of the stress control. Maintenance of genome and epigenome stability under stress costs energy, and a limitation of resources under stress may allow this investment only locally. The less drastic effects of heat stress on higher order nuclear architecture in the shoot apical meristem, compared with differentiated tissues (Pecinka et al. 2010), might indicate such preferential protection, which would, in turn, reduce the chance for trans-generational chromatin changes even more. However, selection on the evolutionary scale, especially under adverse conditions, would certainly favor adaptive changes on all levels, including chromatin, even if they occur only with

minimal probability. Currently, they are not unambiguously substantiated, but plants are good candidates for a further, unprepossessed search. Constant refinement of chromatin analysis tools and growing genomic information, also for non-model species, together with the criteria listed here, will help answer whether it is time for a renaissance of Lamarck's ideas.

## References

Adrian, J., Torti, S. and Turck, F. (2009) From decision to commitment: the molecular memory of flowering. *Mol. Plant* 2: 628–642.

Ahuja, I., de Vos, R.C.H., Bones, A.M. and Hall, R.D. (2010) Plant molecular stress responses face climate change. *Trends Plant Sci.* 15: 664–674.

Becker, C., Hagmann, J., Mueller, J., Koenig, D., Stegle, O., Borgwardt, K. et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* 480: 245–249.

Bilichak, A., Ilnystkyy, Y., Hollunder, J. and Kovalchuk, I. (2012) The progeny of Arabidopsis thaliana plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. *PLoS One* 7: e30515.

Boyko, A. and Kovalchuk, I. (2011) Genome instability and epigenetic modification—heritable responses to environmental stress? *Curr. Opin. Plant Biol.* 14: 260–266.

Butterbrodt, T., Thurow, C. and Gatz, C. (2006) Chromatin immuno-precipitation analysis of the tobacco PR-1a- and the truncated CaMV 35S promoter reveals differences in salicylic acid-dependent TGA factor binding and histone acetylation. *Plant Mol. Biol.* 61: 665–674.

Chandler, V.L. and Stam, M. (2004) Chromatin conversations: mechanisms and implications of paramutation. *Nat. Rev. Genet.* 5: 532–544.

Chinnusamy, V., Schumaker, K. and Zhu, J.K. (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *J. Exp. Bot.* 55: 225–236.

Chinnusamy, V. and Zhu, J.K. (2009) Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* 12: 133–139.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y. et al. (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* 466: 388–392.

Conrath, U. (2011) Molecular aspects of defence priming. *Trends Plant Sci.* 16: 524–531.

Cubas, P., Vincent, C. and Coen, E. (1999) An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157–161.

Daxinger, L. and Whitelaw, E. (2012) Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat. Rev. Genet.* 13: 153–162.

Dhawan, R., Luo, H., Foerster, A.M., AbuQamar, S., Du, H.-N., Briggs, S.D. et al. (2009) HISTONE MONOUBIQUITINATION1 interacts with a subunit of the mediator complex and regulates defense against necrotrophic fungal pathogens in Arabidopsis. *Plant Cell* 21: 1000–1019.

Elmayan, T., Proux, F. and Vaucheret, H. (2005) Arabidopsis RPA2: a genetic link among transcriptional gene silencing, DNA repair, and DNA replication. *Curr. Biol.* 15: 1919–1925.

Foerster, A.M., Dinh, H.Q., Sedman, L., Wohlrab, B. and Mittelsten Scheid, O. (2011) Genetic rearrangements can modify chromatin features at epialleles. *PLoS Genet.* 7: e1002331.

Grativol, C., Hemerly, A.S. and Ferreira, P.C.G. (2012) Genetic and epigenetic regulation of stress responses in natural plant populations. *Biochim. Biophys. Acta* 1819: 176–185.

Guerrero-Bosagna, C. and Skinner, M.K. (2012) Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. *Mol. Cell. Endocrinol.* 354: 3–8.

Gupta, R., Nagarajan, A. and Wajapeyee, N. (2010) Advances in genome-wide DNA methylation analysis. *Biotechniques* 49: III–XI.

Huang, G.-T., Ma, S.-L., Bai, L.-P., Zhang, L., Ma, H., Jia, P. et al. (2012) Signal transduction during cold, salt, and drought stresses in plants. *Mol. Biol. Rep.* 39: 969–987.

Ito, H., Bucher, E., Herve, G., Mirouze, M., Vaillant, I. and Paszkowski, J. (2010) siRNA pathway control transgenerational transposition of heat-activated retrotransposon in Arabidopsis. *Genes Genetic Syst.* 85: 435–435.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. and Paszkowski, J. (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472: 115–119.

Jablonka, E. and Raz, G. (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* 84: 131–176.

Jarillo, J.A., Pineiro, M., Cubas, P. and Martinez-Zapater, J.M. (2009) Chromatin remodeling in plant development. *Int. J. Dev. Biol.* 53: 1581–1596.

Jaskiewicz, M., Conrath, U. and Peterhaensel, C. (2011) Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. *EMBO Rep.* 12: 50–55.

Kanno, T. and Habu, Y. (2011) siRNA-mediated chromatin maintenance and its function in Arabidopsis thaliana. *Biochim. Biophys. Acta* 1809: 444–451.

Khraiwesh, B., Zhu, J.K. and Zhu, J. (2012) Role of miRNA and siRNA in biotic and abiotic stress responses in plants. *Biochim. Biophys. Acta* 1819: 137–148.

Kim, D.-H., Doyle, M.R., Sung, S. and Amasino, R.M. (2009) Vernalization: winter and the timing of flowering in plants. *Annu. Rev. Cell Dev. Biol.* 25: 277–299.

Kim, J.-M., To, T.K. and Seki, M. (2012) An epigenetic integrator: new insights into genome regulation, environmental stress responses and developmental controls by HISTONE DEACETYLASE 6. *Plant Cell Physiol.* 53: 794–800.

Koonin, E.V. and Wolf, Y.I. (2009) Is evolution Darwinian or/and Lamarckian? *Biol. Direct* 4.

Kumar, S.V. and Wigge, P.A. (2010) H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. *Cell* 140: 136–147.

Lang-Mladek, C., Popova, O., Kiok, K., Berlinger, M., Rakic, B., Aufsatz, W. et al. (2010) Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in Arabidopsis. *Mol. Plant* 3: 594–602.

Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11: 204–220.

Lienert, F., Wirbelauer, C., Som, I, Dean, A., Mohn, F. and Schuebeler, D. (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.* 43: 1091–1097.

Luna, E., Bruce, T.J., Roberts, M.R., Flors, V. and Ton, J. (2012) Next generation systemic acquired resistance. *Plant Physiol.* 158: 844–853.

Luo, M., Liu, X., Singh, P., Cui, Y., Zimmerli, L. and Wu, K. (2011) Chromatin modifications and remodeling in plant abiotic stress response. *Biochim. Biophys. Acta* 1819: 129–136.

Ma, K.-W., Flores, C. and Ma, W. (2011) Chromatin configuration as a battlefield in plant–bacteria interactions. *Plant Physiol.* 157: 535–543.

Madlung, A. and Comai, L. (2004) The effect of stress on genome regulation and structure. *Ann. Bot.* 94: 481–495.

Manning, K., Tor, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J. et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* 38: 948–952.

March-Diaz, R., Garcia-Dominguez, M., Lozano-Juste, J., Leon, J., Florencio, F.J. and Reyes, J.C. (2008) Histone H2A.Z and homologues of components of the SWR1 complex are required to control immunity in Arabidopsis. *Plant J.* 53: 475–487.

Matsunaga, W., Kobayashi, A., Kato, A. and Ito, H. (2012) The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of *ONSEN*, a *copia*-like retrotransposon in *Arabidopsis thaliana*. *Plant Cell Physiol.* 53: 824–833.

Meyer, P. (2011) DNA methylation systems and targets in plants. *FEBS Lett.* 585: 2008–2015.

Miao, V.P.W., Freitag, M. and Selker, E.U. (2000) Short TpA-rich segments of the zeta–eta region induce DNA methylation in Neurospora crassa. *J. Mol. Biol.* 300: 249–273.

Mirouze, M. and Paszkowski, J. (2011) Epigenetic contribution to stress adaptation in plants. *Curr. Opin. Plant Biol.* 14: 267–274.

Mosher, R.A., Durrant, W.E., Wang, D., Song, J. and Dong, X. (2006) A comprehensive structure–function analysis of Arabidopsis SNI1 defines essential regions and transcriptional repressor activity. *Plant Cell* 18: 1750–1765.

Palma, K., Thorgrimsen, S., Malinovsky, F.G., Fiil, B.K., Nielsen, H.B., Brodersen, P. et al. (2010) Autoimmunity in Arabidopsis acd11 is mediated by epigenetic regulation of an immune receptor. *PLoS Pathog* 6: e1001137.

Paszkowski, J. and Grossniklaus, U. (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. *Curr. Opin. Plant Biol.* 14: 195–203.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N. and Mittelsten Scheid, O. (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell* 22: 3118–3129.

Pecinka, A., Rosa, M., Schikora, A., Berlinger, M., Hirt, H., Luschnig, C. et al. (2009) Transgenerational stress memory is not a general response in Arabidopsis. *PLoS One* 4: e5202.

Rasmann, S., De Vos, M., Casteel, C.L., Tian, D., Halitschke, R., Sun, J.Y. et al. (2012) Herbivory in the previous generation primes plants for enhanced insect resistance. *Plant Physiol.* 158: 854–863.

Santos, A.P., Serra, T., Figueiredo, D.D., Barros, P., Lourenço, T., Chander, S. et al. (2011) Transcription regulation of abiotic stress responses in rice: a combined action of transcription factors and epigenetic mechanisms. *OMICS* 15: 839–857.

Saze, H. and Kakutani, T. (2011) Differentiation of epigenetic modifications between transposons and genes. *Curr. Opin. Plant Biol.* 14: 81–87.

Saze, H., Tsugane, K., Kanno, T. and Nishimura, T. (2012) DNA methylation in plants: relationship with small RNAs and histone modifications, and functions in transposon inactivation. *Plant Cell Physiol.* 53: 766–784.

Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Urich, M.A., Libiger, O. et al. (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334: 369–373.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastroem, A., Field, Y., Moore, I.K. et al. (2006) A genomic code for nucleosome positioning. *Nature* 442: 772–778.

Seong, K.-H., Li, D., Shimizu, H., Nakamura, R. and Ishii, S. (2011) Inheritance of stress-induced, ATF-2-dependent epigenetic change. *Cell* 145: 1049–1061.

Shah, J. (2009) Plants under attack: systemic signals in defence. *Curr. Opin. Plant Biol.* 12: 459–464.

Sheldon, C.C., Hills, M.J., Lister, C., Dean, C., Dennis, E.S. and Peacock, W.J. (2008) Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization. *Proc. Natl Acad. Sci. USA* 105: 2214–2219.

Skinner, M.K. (2011) Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. *Epigenetics* 6: 838–842.

Slaughter, A., Daniel, X., Flors, V., Luna, E., Hohn, B. and Mauch-Mani, B. (2012) Descendants of primed Arabidopsis plants exhibit resistance to biotic stress. *Plant Physiol.* 158: 835–843.

Soppe, W.J.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M. et al. (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* 6: 791–802.

Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Mittelsten Scheid, O. and Paszkowski, J. (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. *Plant Cell* 12: 1165–1178.

Steward, N., Kusano, T. and Sano, H. (2000) Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. *Nucleic Acids Res.* 28: 3250–3259.

Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C. et al. (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* 323: 1600–1604.

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J. and Vaillant, I. (2010) Stress-induced activation of heterochromatic transcription. *PLoS Genet.* 6: e1001175.

Vaillant, I. and Paszkowski, J. (2007) Role of histone and DNA methylation in gene regulation. *Curr. Opin. Plant Biol.* 10: 528–533.

van den Burg, H.A. and Takken, F.L.W. (2009) Does chromatin re-modeling mark systemic acquired resistance?. *Trends Plant Sci.* 14: 415–415.

Verdeil, J.-L., Alemanno, L., Niemenak, N. and Tranbarger, T.J. (2007) Pluripotent versus totipotent plant stem cells: dependence versus autonomy? *Trends Plant Sci.* 12: 245–252.

Walley, J.W., Rowe, H.C., Xiao, Y., Chehab, E.W., Kliebenstein, D.J., Wagner, D. et al. (2008) The chromatin remodeler SPLAYED regulates specific stress signaling pathways. *PLoS Pathog* 4: e100237.

Yaish, M.W., Colasanti, J. and Rothstein, S.J. (2011) The role of epigen-etic processes in controlling flowering time in plants exposed to stress. *J. Exp. Bot.* 62: 3727–3735.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* 57: 781–803.

Zhu, Y., Dong, A.W. and Shen, W.H. (2012) Histone variants and chro-matin assembly in plant abiotic stress response. *Biochim. Biophys. Acta* 1819: 343–348.

Zilberman, D., Coleman-Derr, D., Ballinger, T. and Henikoff, S. (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* 456: 125–129.

# Publikace 11

# Cooperation of Multiple Chromatin Modifications Can Generate Unanticipated Stability of Epigenetic States in *Arabidopsis* [C][W][OA]

Tuncay Baubec,[a,1] Huy Q. Dinh,[a,b] Ales Pecinka,[a] Branislava Rakic,[a] Wilfried Rozhon,[a,2] Bonnie Wohlrab,[a] Arndt von Haeseler,[b] and Ortrun Mittelsten Scheid[a,3]

[a] Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, 1030 Vienna, Austria
[b] Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, 1030 Vienna, Austria

Epigenetic changes of gene expression can potentially be reversed by developmental programs, genetic manipulation, or pharmacological interference. However, a case of transcriptional gene silencing, originally observed in tetraploid *Arabidopsis thaliana* plants, created an epiallele resistant to many mutations or inhibitor treatments that activate many other suppressed genes. This raised the question about the molecular basis of this extreme stability. A combination of forward and reverse genetics and drug application provides evidence for an epigenetic double lock that is only alleviated upon the simultaneous removal of both DNA methylation and histone methylation. Therefore, the cooperation of multiple chromatin modifications can generate unanticipated stability of epigenetic states and contributes to heritable diversity of gene expression patterns.

## INTRODUCTION

Genetically determined loss of gene expression by mutation, insertion of transposons, or chromosomal rearrangements is usually irreversible, since the chance of precisely reconstituting the original DNA sequence is low. On the other hand, epigenetic loss of gene activity is defined as not affecting the DNA sequence but rather as chemically modifying DNA and associated proteins, thus altering the packaging of chromatin and its accessibility for the transcription machinery. Affected sequences are kept transcriptionally inactive by well-characterized pathways that establish DNA methylation and/or histone modifications. For several of these modifications, antagonistic enzymes have been described (Chen and Tian, 2007; Pfluger and Wagner, 2007; Ooi and Bestor, 2008), and many epigenetically regulated sequences undergo a cycle of silencing and activation in the life cycle of the organism. Familiar examples in developmental programs are imprinted genes, dosage-compensated chromosomes, or master regulatory genes under the control of the Polycomb/Trithorax system. Even genetic templates that can produce potentially

deleterious transcripts and are usually under tight epigenetic control can become activated under stress conditions (for review, see Madlung and Comai, 2004; Chinnusamy and Zhu, 2009) or in the germ line in order to reinforce silencing via small RNA during transmission of genetic material to the next generation (Brennecke et al., 2008; Slotkin et al., 2009). However, some cases of genes with very durable epigenetic marks are also known (Chong and Whitelaw, 2004), and the stable transmission of their epigenetic state to subsequent generations has led to their denotation as epialleles (Finnegan, 2002). Examples in plants are a methylated transcription factor gene changing flower morphology in *Linaria* (Cubas et al., 1999) and the pigmentation-controlling transcription factor genes in maize (*Zea mays*) that are downregulated by paramutation (for review, see Chandler et al., 2000). These famous cases were identified because of the striking phenotypes. It is likely that many more epialleles exist with less drastic morphological consequences but which nevertheless make a significant contribution to natural evolution and plant breeding (Kalisz and Purugganan, 2004).

Epialleles with remarkable stability have been observed in various tetraploid lines of *Arabidopsis thaliana* derived from a common diploid progenitor (Mittelsten Scheid et al., 2003). The transgenic resistance marker gene, hygromycin phosphotransferase (*HPT*), under the control of the strong, constitutively active promoter of the *Cauliflower mosaic virus* (P35S) was present in these genetically identical lines either in fully active or completely silenced state. Both states were maintained during backcrosses to diploid lines homozygous for the *HPT*, giving rise to diploid lines C2R (resistant to hygromycin, active *HPT*) and C2S1 (diploid, sensitive to hygromycin, silent *HPT*). In crossing experiments with the tetraploid lines, the epialleles exerted a paramutation-like interaction in which the silent epiallele led to inactivation of the previously active counterpart (Mittelsten

Scheid et al., 2003). The epialleles differ in the degree of DNA methylation and histone modification patterns (Hetzl et al., 2007; Foerster, 2009), as do many other active and inactive sequences. The epialleles show an extremely tight silencing (as described in the following): they were originally found in the tetraploid lines, and the epiallelic interaction occurred only in tetraploid inter-crosses. Therefore, we refer to this phenomenon as polyploidy-associated transcriptional gene silencing (paTGS) even in the diploid lines. Most higher plants are polyploid (Masterson, 1994), and polyploidy is assumed to be a very important driving force in plant evolution and breeding (Stebbins, 1966). Furthermore, epigenetic changes are frequent in freshly formed polyploids (for review, see Osborn et al., 2003; Adams and Wendel, 2005). Paramutation-like epiallelic interaction can lead to significant shifts in the distribution of traits within populations of polyploid plants and drive their evolution more rapidly than anticipated by classical Mendelian genetics. Therefore, it is important to un-derstand the characteristics of the epialleles that underwent paTGS. The described silent *HPT* epiallele offered an excellent model for this analysis, since its stability also in the diploid derivative line and the encoded protein allowed a selection-based genetic screen for *trans*- and *cis*-acting factors involved in the maintenance of the silencing. Here, we demonstrate that the silent epiallele derived from the tetraploid line is under a double safeguard mechanism, which requires the concomitant loss of methylation of both DNA and histones for restoration of tran-scription. This is in contrast with many other transcriptionally silent sequences in the *Arabidopsis* genome that can be acti-vated by removing only one of several inactive chromatin marks by mutation or specific inhibitors. Thus, epialleles in polyploid plants cannot easily revert and represent particularly stable states that are under tight control. For this reason, they might be highly relevant for long-term adaptation of gene expression patterns, breeding, and natural evolution.

## RESULTS

### paTGS Is Resistant to Treatments with DNA Methylation and Histone Deacetylase Inhibitors

Transcriptional inactivation in plants and mammals is frequently associated with methylation of cytosine residues in the DNA, an exchange of specific methylation of histone tails from active to inactive marks, and general deacetylation of histone tails (Chen and Tian, 2007; Vaillant and Paszkowski, 2007). Inhibitors spe-cific for DNA methyltransferases and histone deacetylases exist, and they have been widely used as potentially activating agents for epigenetically silenced endogenes and transgenes (Chang and Pikaard, 2005). The DNA methylation inhibitor zebularine (ZEB) (Zhou et al., 2002) and the histone deacetylase inhibitor trichostatin A (TSA) (Yoshida et al., 1995) were therefore applied to test whether they would reactivate the silent *HPT* transgene. Seeds from the diploid line C2S1 with the inactive *HPT* and seeds from the *HPT*-expressing, hygromycin-resistant line C2R were germinated and plantlets grown for 3 weeks on plates containing 10 μg/mL of hygromycin in combination with 40 μM ZEB and/or 1.6 μM TSA, concentrations that were previously described to be

effective in reactivating silenced targets and reducing methyla-tion in all possible sequence contexts (Baubec et al., 2009) or were even higher than effective concentrations (Chang and Pikaard, 2005). ZEB causes growth retardation but allows the *HPT*-expressing line C2R to grow under selection upon all treatments. By contrast, no growth was observed in line C2S1 (Figure 1A), even upon sequential application of the drugs prior to selection. The applied drug treatments could not, therefore, reactivate the HPT gene and restore the resistant phenotype.

Stringent hygromycin selection requires a certain amount of *HPT* RNA and protein to be produced. To determine whether the inhibitors would release subthreshold levels of gene expression, we performed RNA gel blot analysis using *HPT*-specific probes on total RNA extracted from C2S1 seedlings treated with 0, 20, 40, and 80 μM ZEB. These showed a minimal increase in *HPT* transcript but substantially less hybridization signal than in C2R (Figure 1B). In addition, known epigenetic mutations, such as *cmt3*, *drm1,2*, and *kyp* that could not restore hygromycin resis-tance after introgression of the silent C2S1 epiallele (Milos, 2006), did also not further enhance the effect of zebularine treatments (see Supplemental Figure 1 online). Surprisingly, RNA gel blot analysis with a probe for a noncoding RNA transcribed from another copy of the P35S promoter, downstream of and in close proximity to the *HPT* gene (see Supplemental Figure 2 online), revealed strong reactivation of this second transcript after ZEB treatment of C2S1 (Figure 1B). The pharmacological demethyl-ation was effective, as demonstrated by methylation-sensitive restriction digest and subsequent DNA gel blotting (Figure 1C), but was not sufficient to reactivate the *HPT*-driving promoter.

### paTGS Can Be Released by Novel *DDM1* and *HOG1* Mutant Alleles

Since the silent *HPT* transgene allowed for a reactivation assay based on hygromycin selection, we performed a forward genetic screen to identify factors involved in this robust epigenetic regulation of the *HPT* promoter. Diploid C2S1 plants carrying the silent *HPT* transgenic locus were mutagenized by random T-DNA insertion, and M2 progeny of 20,000 independent trans-formants was screened for hygromycin resistance. We identified three novel alleles of *DECREASE IN DNA METHYLATION1* (*DDM1*), a member of the ATP-dependent SWI2/SNF2 chromatin remodeling gene family (Vongs et al., 1993; Mittelsten Scheid et al., 1998; Jeddeloh et al., 1999) and one novel allele of the *HOMOLOGOUS GENE SILENCING1* (*HOG1*) gene, coding for an *S*-adenosyl-L-homocysteine (SAH) hydrolase (SAHH) (Furner et al., 1998; Rocha et al., 2005). Mutations in *DDM1* (At5g66750) have been previously shown to interfere with main-tenance of transcriptional gene silencing at numerous endoge-nous and transgenic inserts by decreasing DNA and H3K9 methylation (Mittelsten Scheid et al., 1998; Jeddeloh et al., 1999; Soppe et al., 2002; Mathieu et al., 2003). HOG1 (or SAHH1, At4g13940) is required to convert SAH into homocysteine. This degradation is essential for recycling of the methyl-group donor *S*-adenosyl-L-methionine (SAM) and prevents inhibition of *trans*-methylation reactions through increased levels of SAH (Weretilnyk et al., 2001). HOG1 is involved in maintaining tran-scriptional gene silencing at numerous targets (Furner et al., 1998;
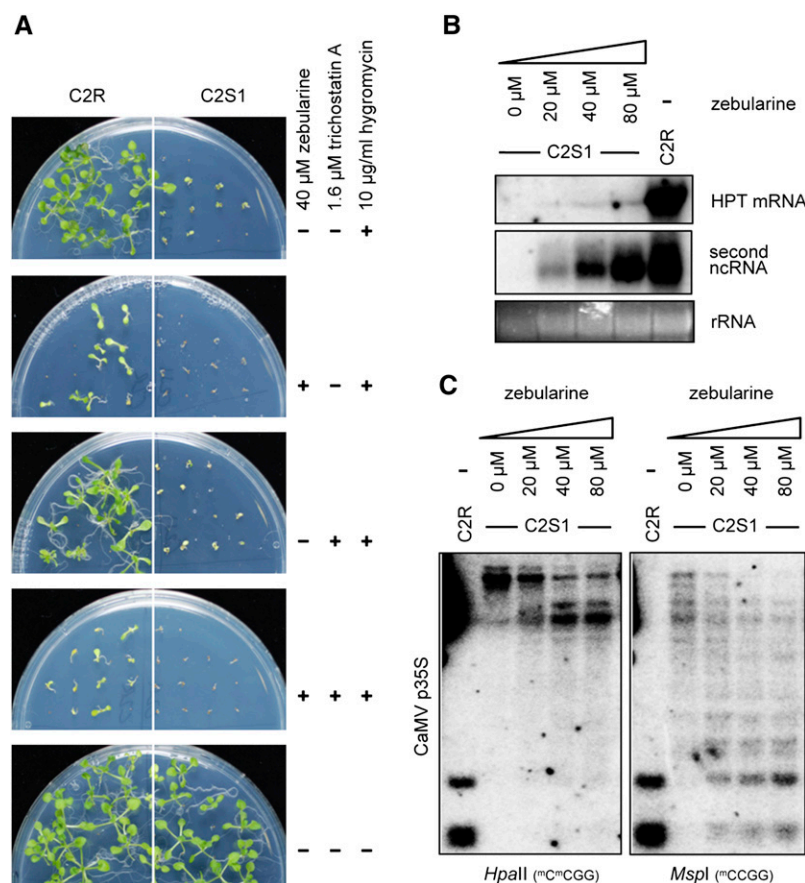
**Figure 1.** Treatments with DNA Methylation and Histone Deacetylase Inhibitors Do Not Release paTGS Silencing.

**(A)** C2R and C2S1 seedlings grown on 10 µg/mL hygromycin plates in the presence of 40 µM ZEB and/or 1.6 µM TSA.
**(B)** RNA gel blot analysis indicates reactivation of the second noncoding transcript but not *HPT* mRNA after 20, 40, and 80 µM ZEB treatments.
**(C)** DNA gel blot analysis of DNA methylation after treatments with increasing zebularine concentrations using promoter-specific probes.
[See online article for color version of this figure.]

Rocha et al., 2005; Mull et al., 2006; Jordan et al., 2007), while another SAHH-related gene (*SAHH2*, At3g23810) has no role in silencing or DNA methylation (Rocha et al., 2005). The *DDM1* alleles were named *ddm1-11* to *ddm1-13*, in continuation of the already available mutant alleles (Jeddeloh et al., 1999; Jordan et al., 2007): *ddm1-11* has a 38-bp deletion in exon V, *ddm1-12* has a 30-bp deletion in exon XIV, and *ddm1-13* has a T-DNA insertion in exon VII (Figure 2A). In contrast with the widely used alleles *ddm1-2*, with a point mutation generating a G-to-A transition in the splice donor site of intron XI (Jeddeloh et al., 1999), and *ddm1-5*, with an 82-bp insert in exon II (Mittelsten Scheid et al., 1998; Jeddeloh et al., 1999), the new mutations are all in conserved signature motifs that are characteristic of SWI2/SNF2 family proteins (Bork and Koonin, 1993) and affect the domains that are important for ATP-dependent chromatin remodeling, namely, SNF2_N and DEAD/DEAH (Figure 2A). This may explain why plants with the new alleles survived the stringent hygromycin selection in the M2 generation during the screen, while plants with the *ddm1-5* allele showed partial reactivation and survived only in F4 after introgression (Mittelsten Scheid et al., 2003; Milos,

2006). A direct comparison of the *ddm1-5* F4 seedlings with the corresponding M4 generation seedlings obtained from the novel alleles further illustrates the differences in resistance (Figure 2B), confirmed by *HPT* expression analysis (see below). *ddm1-12* was used as a representative *ddm1* allele in the following experiments. The new *HOG1* allele, named *hog1-7* in continuation of previously identified alleles (Rocha et al., 2005), has a rearranged T-DNA insertion in the 3′ UTR (Figure 2C). Although this mutation is not likely to cause a complete loss of function, it affects *HOG1* mRNA levels and stability, as revealed by quantitative RT-PCR (Figure 2D).

We analyzed the degree of *HPT* reactivation in 3-week-old M4 mutant seedlings. Quantification of *HPT* transcripts with real-time PCR using cDNA obtained from reverse-transcribed total RNA from the *ddm1-12* and *hog1-7* mutants indicated a similar abundance as in the active line C2R (1-fold $\pm$ 0.35 and 0.96-fold $\pm$ 0.23 in *hog1-7* and *ddm1-12* mutants, respectively; Figure 3A).

This is in agreement with a similar loss of DNA methylation at the P35S promoter, as shown by DNA gel blot analysis (Figure 3B). To quantify the degree of DNA demethylation specifically at the promoter upstream of *HPT*, we applied bisulfite sequencing
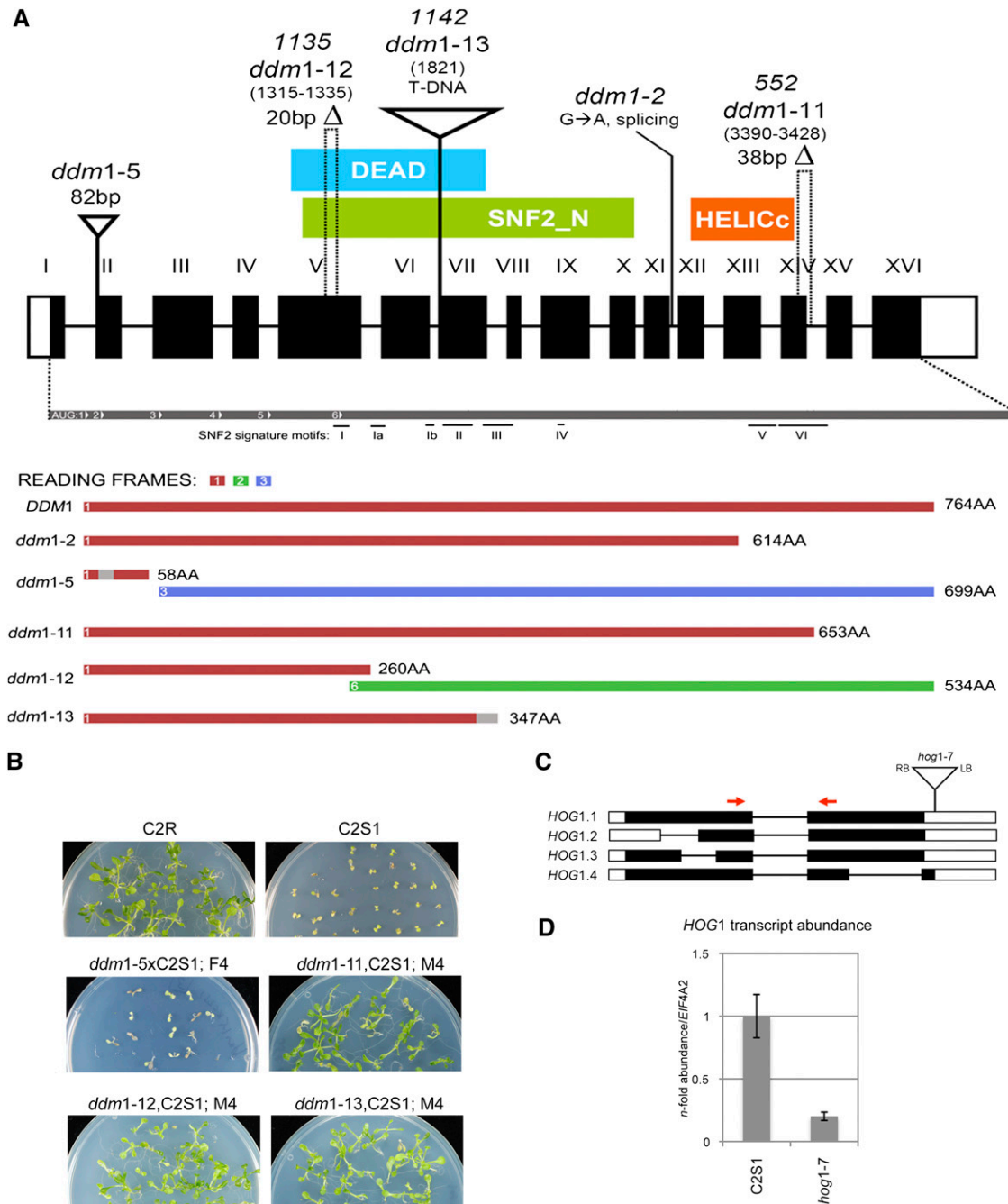
**Figure 2.** Novel *ddm1* and *hog1* Mutant Alleles.

**(A)** *DDM1* gene region with indicated UTRs (white boxes), exons (filled boxes), and introns (lines). Functional domains are indicated by colored boxes, while mutations are indicated by insertions or deletions (Δ). Below: reading frame analysis in the *ddm1* alleles. Coding sequence is indicated by the gray bar, and conserved SWI2/SNF2 signatures (Bork and Koonin, 1993) are shown below. White glyphs indicate potential translation initiation sites in the 5′ region (aa(A/G)(A/C)a**AUG**Ccg; Rangan et al., 2008). Coding reading frames (in different colors) and encoded protein size are predicted in wild-type and mutant alleles. Light-gray bars indicate nonplant DNA insertions.
**(B)** Allele comparison by hygromycin selection in analogous generations: F4 from crosses between C2S1 × *ddm1-5* and M4 in the novel alleles. C2S1 and C2R are used as controls.
**(C)** Mutant integration site in the *SAHH/HOG1* gene. UTRs are indicated as white boxes, exons as filled boxes, and introns as lines. The four predicted splice variants are displayed (TAIR7).
**(D)** Quantification of *HOG* transcript abundance in wild-type C2S1 and *hog1-7* mutant plants normalized to *EIF4A2*. Error bars represent SD from triplicate analyses. Used primers are indicated by red arrows in **(C)**.
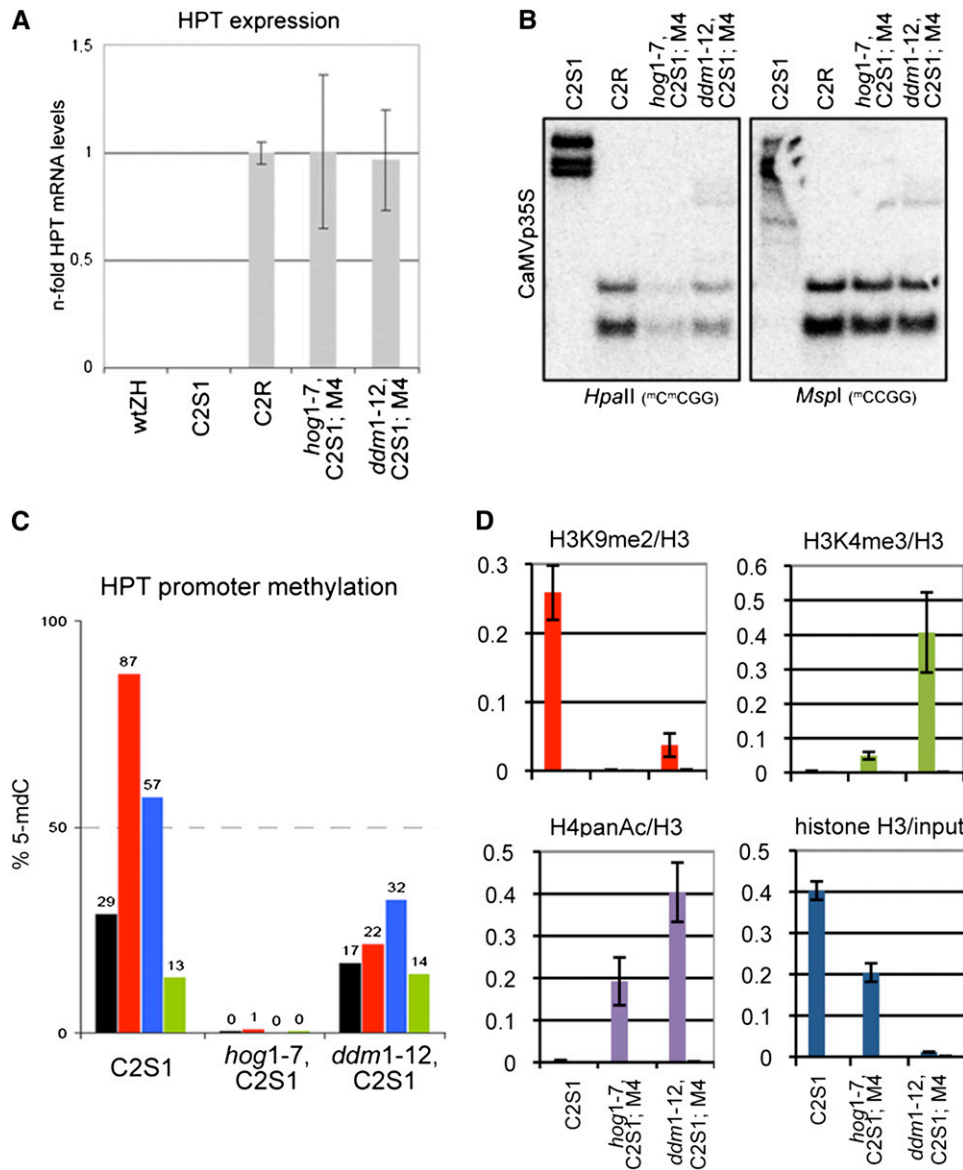
**Figure 3.** Mutations in *DDM1* and *HOG1* Release paTGS from the HPT Transgene.

**(A)** Quantification of HPT mRNA levels in wild-type, C2S1, C2R *hog1-7*,C2S1, and *ddm1-12*,C2S1 seedlings normalized to *EIF4A2*. Error bars represent SD from triplicate analyses.
**(B)** p35S DNA methylation analysis in C2S1, C2R, and mutant plants by DNA gel blotting of DNA digested by methylation-sensitive restriction enzymes.
**(C)** Promoter DNA methylation analysis by bisulfite sequencing representing total (black) and sequence context-specific (ᵐCG, red; ᵐCHG, blue; ᵐCHH, green) methylation.
**(D)** Analysis of histone modifications and histone H3 abundance normalized to H3 or input at the *HPT* promoter by ChIP in C2S1 and mutant lines. Gray columns (right of the colored columns and very small) represent samples precipitated without antibodies.

to DNA of the *hog1-7* and *ddm1-12* mutants (Figure 3C). Total DNA methylation was reduced from 29% in C2S1 to 0% in *hog1-7* and to 17% in *ddm1-12*, while CG-specific methylation was reduced from 87 to 1% and 22% in *hog1-7* and *ddm1-12*, respectively. We observed a similar decrease in methylation at CHG sites, where the *hog1-7* mutation resulted in 0% residual methylated CHG sites, while the *ddm1-12* mutation maintained 32% of the methylated CHG sites compared with 57% in C2S1.

CHH-specific methylation, with 13% of all available sites in C2S1, was significantly decreased in *hog1-7* with 0%, while it remained unaltered in the *ddm1-12* mutant, indicating that DDM1 is not required to maintain methylation at these largely nonsymmetrical sites (Figure 3C).

To complement the analysis of chromatin changes in the mutants, we further analyzed *ddm1-12*–specific and *hog1-7*–specific changes in histone modifications at the *HPT* transgene promoter

by chromatin immunoprecipitation (ChIP) using specific antibodies or antisera against the heterochromatic mark histone H3 Lys-9 dimethylation (H3K9me2) and the euchromatic marks histone H3 Lys-4 trimethylation (H3K4me3) and histone H4 panacetylation (H4panAc) (for review, see Fuchs et al., 2006). General nucleosome occupancy in the examined regions was analyzed by ChIP with antibodies recognizing histone H3 independent of modifications. Enrichment of the DNA fragments in the modification-specific precipitates was measured by quantitative PCR in triplicate and was related to their loading with histone H3. The prevalence of heterochromatic H3K9me2 in C2S1 was drastically reduced in the *hog1-7* and *ddm1-12* mutants (Figure 3D, red columns). H3K4me3 increased in both mutants compared with C2S1, although only slightly in *hog1-7* and much more pronounced in *ddm1-12* (Figure 3D, green columns). An increase of H4 acetylation was observed in both mutants, again with a stronger increase in *ddm1-12* (Figure 3D, violet columns). Remarkably, nucleosome occupancy measured as histone H3 abundance relative to input DNA was comparable between C2S1 and *hog1-7* but almost totally lost in *ddm1-12*. This should be considered when interpreting the relative enrichment or depletion of histone marks in the mutants (Figure 3D, blue columns).

## Mutations in *DDM1* or *HOG1* Affect Methylation of DNA and Histones Globally

Transcriptional silencing associated with DNA methylation and heterochromatic marks can be released by different means, including specific inhibitors or loss of function of epigenetic regulators. As shown above, the silent *HPT* transgene that was found in the polyploid lines did not respond to inhibitors. It also remained suppressed in the background of many mutations representing the known epigenetic regulatory pathways (Milos, 2006; Baubec et al., 2009; Foerster, 2009). This raised the question of why and how the new mutations in *DDM1* and *HOG1* proved to be exceptions and whether this could hint at an underlying mechanism. Both mutants have been reported to interfere with transcriptional gene silencing at many other targets in the *Arabidopsis* genome (Lippman et al., 2004; Jordan et al., 2007), but many of these were also expressed in those other mutants that did not reactivate the *HPT* gene. However, mutations in *DDM1* and *HOG1* have in common that they reduce DNA methylation and heterochromatic histone modifications at the *HPT* transgene. This effect of *DDM1* loss has also been described for other targets (Gendrel et al., 2002; Johnson et al., 2002; Soppe et al., 2002; Probst et al., 2003). Mutations in *HOG1* cause DNA hypomethylation at transgenic and endogenous repeats (Furner et al., 1998; Rocha et al., 2005; Mull et al., 2006), and the function of the *HOG1* gene regulating the level of the methyl group donor SAM suggested that its loss would also affect histone methylation (Rocha et al., 2005). To challenge the hypothesis that removal of both marks is a common feature of *ddm1* and *hog1*, we characterized DNA methylation and histone methylation in the novel mutant alleles in general and also at other sequences to allow for a direct comparison of the extent and specificity of the effects.

In agreement with results published for other alleles (Vongs et al., 1993; Furner et al., 1998; Jeddeloh et al., 1999; Rocha

et al., 2005), global DNA methylation in *hog1-7* and *ddm1-12* was reduced to 2.7% ($\pm$0.47) and 1.7% ($\pm$0.13), respectively, in comparison to 5-methyldeoxycytosine (5-mdC) levels of 5.9% ($\pm$0.5) in the parental line C2S1, which is similar to wild-type levels (Rozhon et al., 2008) (Figure 4A). A significant proportion of the DNA methylation in wild-type *Arabidopsis* is found at repetitive sequences (Martinez-Zapater et al., 1986) and disappears in *ddm1* or *hog1* mutants (Vongs et al., 1993; Furner et al., 1998). This is also true for the new alleles: DNA gel blot analysis of DNA methylation at centromeric 180-bp repeats (Figure 4B) showed drastic hypomethylation in both mutants. However, the demethylation was more pronounced in the *ddm1-12* mutant, especially for the CG sites (Figure 4B). A certain difference was also evident after cytological analysis of the usually compact heterochromatic chromocentres by immunofluorescence, revealing dispersed 5-mdC localization in *ddm1-12*, where just 14% (n = 104) of nuclei retained chromocentric 5-mdC signals (Figure 4C). This is in agreement with other reports (Soppe et al., 2002). Nuclei of *hog1-7*, however, maintained most 5-mdC (89%, n = 80) at the chromocenters (CCs), close to wild-type nuclei (91%, n = 109), in accordance with the DNA gel blot methylation analysis of the centromeric repeats. This suggests that loss of DNA methylation in the *hog1-7* mutant occurs primarily at other parts of the genome. H3K9me2, as revealed by immunostaining, also colocalizes with CCs in wild-type nuclei (71%, n = 129) but is reduced in both mutants to 8 and 10% of nuclei having wild-type morphology (n = 114 and 107, respectively; Figure 4D), as also previously reported for *ddm1* (Probst et al., 2003).

The loss of chromocentric H3K9me2 signals in *hog1-7* nuclei, independent of the remaining DNA methylation, suggests a direct effect of SAM depletion on histone methylation. The cytological evidence was further substantiated by loss of silencing accompanied by reduced DNA and histone methylation at the retrotransposon without long terminal repeats *LINE1-4* (At2g01840) (Lippman et al., 2003) in the *ddm1-12* and *hog1-7* mutations (see Supplemental Figures 3A to 3C online), as well as by decreased levels of H3K9me2 and H3K4me3 in *hog1-7* analyzed by immunoblot (see Supplemental Figure 3D online). This provides further evidence of globally reduced histone methylation in *hog1* mutants, independent of the Lys residue analyzed.

## Inhibition of SAHH Interferes with Maintenance of paTGS

The similar but not identical consequences of mutations in *DDM1* and *HOG1* on general DNA and histone methylation let us postulate that their comparable and exclusive role among TGS mutants in the maintenance of paTGS would occur through directly and simultaneously affecting DNA and histone methylation at the HPT promoter. A genetic approach to simultaneously reduce histone methylation and DNA methylation in all sequence contexts would require combination of at least six mutations and renders plants with severe developmental aberrations (Chan et al., 2006; Johnson et al., 2008). Therefore, we tried to mimic the *hog1* mutation by applying the specific SAHH inhibitor dihydroxypropyladenine (DHPA). The adenosine homolog DHPA was shown to induce hypomethylation and release of posttranscriptionally silenced transgenes in tobacco (*Nicotiana*
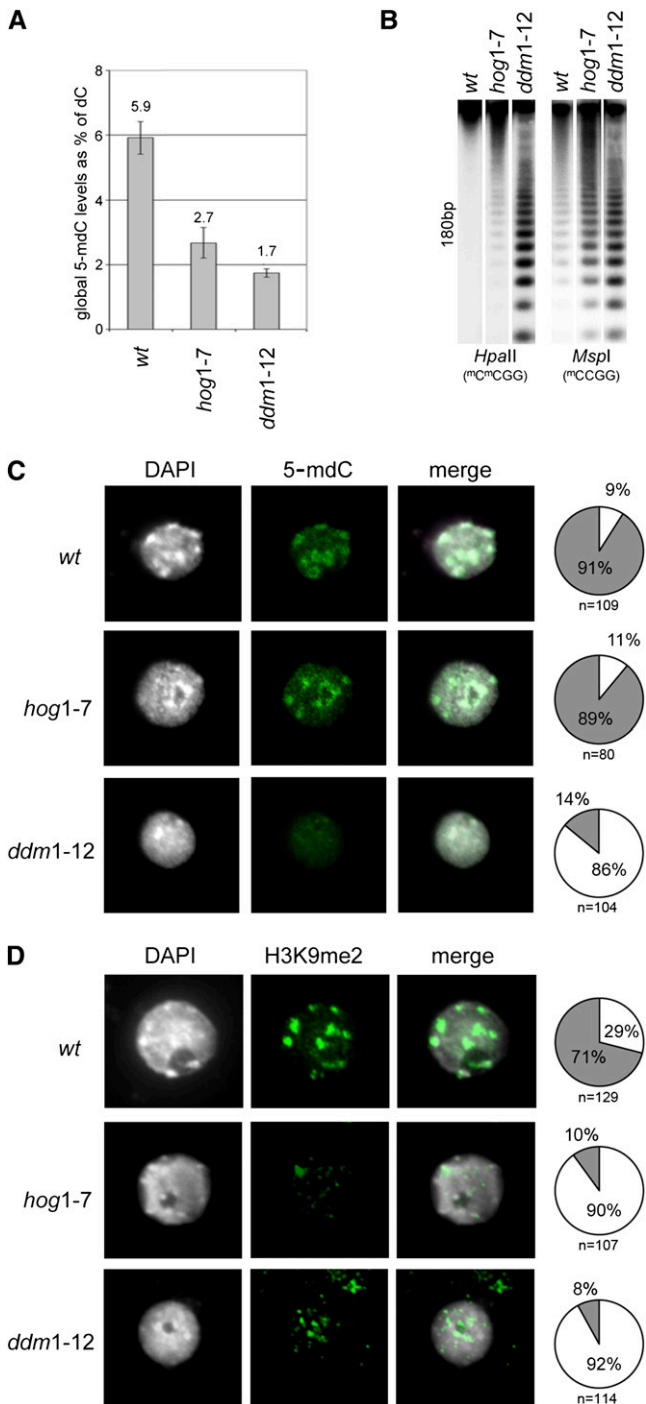
**Figure 4.** Mutations in *HOG1* and *DDM1* Lead to a Global Decrease of DNA and Histone Methylation.

**(A)** Global 5-mdC levels measured by HPLC are reduced in *hog1-7* and *ddm1-12* seedlings.

**(B)** DNA gel blot analysis showing decreased DNA methylation at centromeric 180-bp repeats in *hog1-7* and *ddm1-12* mutant plants.

**(C)** Chromocentric 5-mdC localization measured by immunofluorescence is lost only in *ddm1-12* but not in *hog1-7*. DAPI, 4',6-diamidino-2-phenylindole.

**(D)** H3K9me2 compaction measured by immunofluorescence is disrupted in both mutants. The pie charts represent the percentage of nuclei with corresponding morphology. Gray, compact signals; white, dispersed signals.

*tabacum*; Kovarik et al., 1994, 2000a). We first established the applicable dose range in *Arabidopsis* and analyzed the effectiveness of DHPA by germinating and growing seeds of a line with a transcriptionally silent, highly repetitive β-glucuronidase (GUS) transgene insertion on chromosome III (L5) (Morel et al., 2000) that is reactivated in the background of numerous epigenetic mutations (Elmayan et al., 2005), including *hog1-7* and *ddm1-12*, or by treatment with DNA methylation inhibitors (Baubec et al., 2009). DHPA treatments had only mild growth effects at the applied concentrations of 50 to 200 μM but successfully induced transcriptional reactivation of the GUS transgene (see Supplemental Figures 4A and 4B online). We subsequently compared DHPA inhibitors with drugs that change either DNA or histone modification. There is no inhibitor that specifically reduces histone methylation while leaving DNA methylation undisturbed. Thus, we applied the histone deacetylation inhibitor TSA, which has repeatedly been shown to convert silent into transcriptionally active genes (Chen and Pikaard, 1997; Xu et al., 2005). ZEB interferes specifically with DNA methylation (Zhou et al., 2002). We performed a side-by-side comparison of wild-type seedlings grown for 3 weeks on media containing either TSA, ZEB, or DHPA in the previously established dose ranges (Chang and Pikaard, 2005; Baubec et al., 2009; this article). We first analyzed transcriptional activation of endogenous repeats by quantitative real-time PCR. As observed in the mutant background (see Supplemental Figures 3A and 3B online), the retrotransposon without long terminal repeats *LINE1-4* (At2g01840) showed significant and dose-dependent transcript abundance (Figure 5A) and DNA hypomethylation (Figure 5B) after ZEB or DHPA, but not TSA treatments. Corresponding to the degree of transcriptional activation, we observed a significant, though not complete, reduction of H3K9me2 (Figure 5C). The active mark H3K4me3 increased but did not reach the levels seen in *ddm1* mutants (see Supplemental Figure 3C online). This is plausible since both histone modifications are likely to require SAM, which is a limiting factor in *hog1* and upon DHPA but not in *ddm1*. Data describing expression, DNA methylation, and histone modification for two other genomic sequences and cytological analysis of treated nuclei support the findings (see Supplemental Figures 5 and 6 online).

Although the retroelement *LINE1-4*, other repetitive sequences, and the second promoter of the transgene were transcriptionally activated by ZEB alone, silencing at the *HPT* promoter itself was not released by this drug (Figure 1B). Therefore, we evaluated the effects of DHPA treatments on expression, DNA methylation, and histone modification of the silenced *HPT* gene, asking whether the drugs would mimic the effects of the *hog1* mutation and release paTGS. The answer was affirmative, and high concentrations (200 μM) of SAHH inhibitor resulted in *HPT* expression up to 60% of the level in the hygromycin-resistant line C2R (Figure 6A). RNA gel blots with specific probes
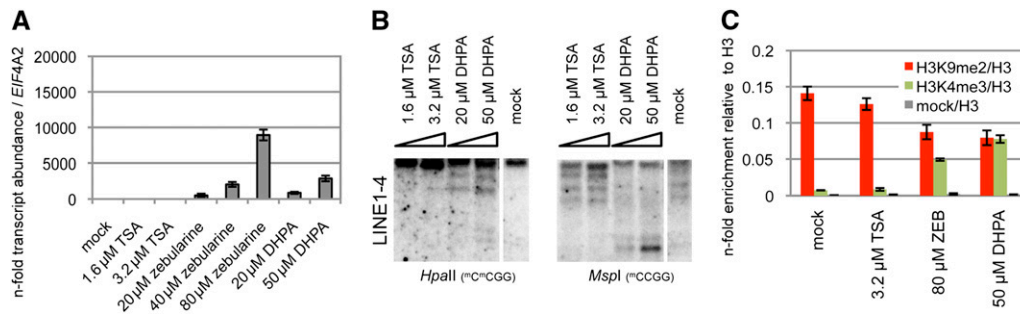
**Figure 5.** The SAHH Inhibitor DHPA Interferes with Transcriptional Gene Silencing at *LINE1-4*.

**(A)** Quantitative RT-PCR measuring the abundance of *LINE1-4* mRNA after chromatin drug and SAHH inhibitor treatments. Error bars denote SD from triplicate analyses.
**(B)** DNA gel blot analysis of DNA methylation with *LINE1-4*–specific probes after chromatin drug and SAHH inhibitor treatments.
**(C)** ChIP analysis of H3K9me2 (red) and H3K4me3 (green) histone modifications after chromatin drug and SAHH inhibitor treatments. Gray columns denote samples precipitated without antibodies.
[See online article for color version of this figure.]

revealed that both promoters were activated concordantly, with the primary promoter producing nearly as much *HPT* transcript as in the active state of the control line C2R (Figure 6B). DNA gel blot analysis of DNA methylation indicated dose-dependent hypomethylation at both promoters upon DHPA treatment, with CHG methylation more affected than CG (Figure 6C). This is in accordance with gradual demethylation at the *Hpa*II/*Msp*I recognition sequence $^mC^mCGG$ sites after DHPA treatments (Kovarik et al., 2000b). Quantification of histone modifications at the *HPT* promoter after DHPA treatments revealed loss of H3K9me2 and a slight gain of H3K4me3 already after 50 $\mu$M DHPA treatments (Figure 6D), as in *hog1-7*. In summary, the chemical interference produced by DHPA application has a similar effect as the genetically determined decrease of functional SAHH by the *hog1-7* mutation. Both cause a reduction of methylation of DNA and the associated histones of several genomic sequences, including the *HPT* transgene that underwent polyploidy-associated gene silencing. The lack of *HPT* reactivation upon depletion of only one type of methylation, in contrast with its restored transcription upon interference with both modifications simultaneously, suggests that this epiallele, and probably similar ones, are under a double-safeguard control that renders gene suppression extremely stable against epigenetic perturbation (Figure 7).

## DISCUSSION

An undisputed definition of epigenetic inheritance is still lacking, but most descriptions refer to its reversible nature to distinguish it from genetic alterations inscribed in the DNA sequence. As is often the case in biology, this sharp distinction does not hold upon closer inspection. While many epigenetically regulated genes undergo programmed, regular, or random cycles of activation and suppression in the course of development, others have proven to be extremely stably silenced. Among them are many transposable elements, for which redundant control by different DNA methyltransferases (Kato et al., 2003) or a special reinforcement by small RNA silencing in the germ line (Brennecke

et al., 2008; Slotkin et al., 2009) have been described. However, even transposons exhibit a surprising diversity in response to epigenetic interference in *Arabidopsis* where the role of well-defined epigenetic pathways can be studied in numerous mutants. Loss-of-function of DNA methyltransferases, argonaute proteins, histone methyltransferases, or histone deacetylases causes transcriptional activation of overlapping but not identical subsets of elements (Lippman et al., 2003). Most of these elements can also be activated by drugs that reduce either DNA methylation or histone modifications (Chang and Pikaard, 2005). Here, we have described a case of epigenetic transcriptional silencing that is surprisingly resistant to genetic and chemical interference, since removing one chromatin modification alone does not restore transcriptional activity from the potentially strong viral P35S promoter. Based on results from forward and reverse mutational screens (Milos, 2006; Baubec, 2008; Foerster, 2009), complemented by inhibitor experiments, we have provided evidence that two epigenetic features, namely, symmetric DNA methylation and histone methylation, cooperate to generate a double safeguard system that controls transcriptional suppression. Hence, both modifications have to be unlocked to convert the silent epiallele into an active one.

This could be achieved by a loss of functional DDM1, a member of the SWI2/SNF2 chromatin remodeling ATP-dependent helicase family. Mutations in *DDM1* known to decrease DNA methylation (Vongs et al., 1993; Kakutani et al., 1996, 1999) also reduce the levels of histone H3 dimethylation at Lys-9 (Gendrel et al., 2002; Habu et al., 2006). A partial interdependence of DNA methylation and H3K9me2 in *Arabidopsis* was further described in mutants of other genes whose products were supposed to act primarily on either DNA or H3K9 methylation (Johnson et al., 2002; Soppe et al., 2002; Tariq et al., 2003). These studies reveal a complex and possibly mutual interplay of DNA and histone methylation at different targets that can also depend on transcriptional activity. However, this interdependence does not apply to the silencing described in this study, since DNA or histone methyltransferase mutations alone did not reactivate the silent epiallele in our study. Even the concomitant reduction of
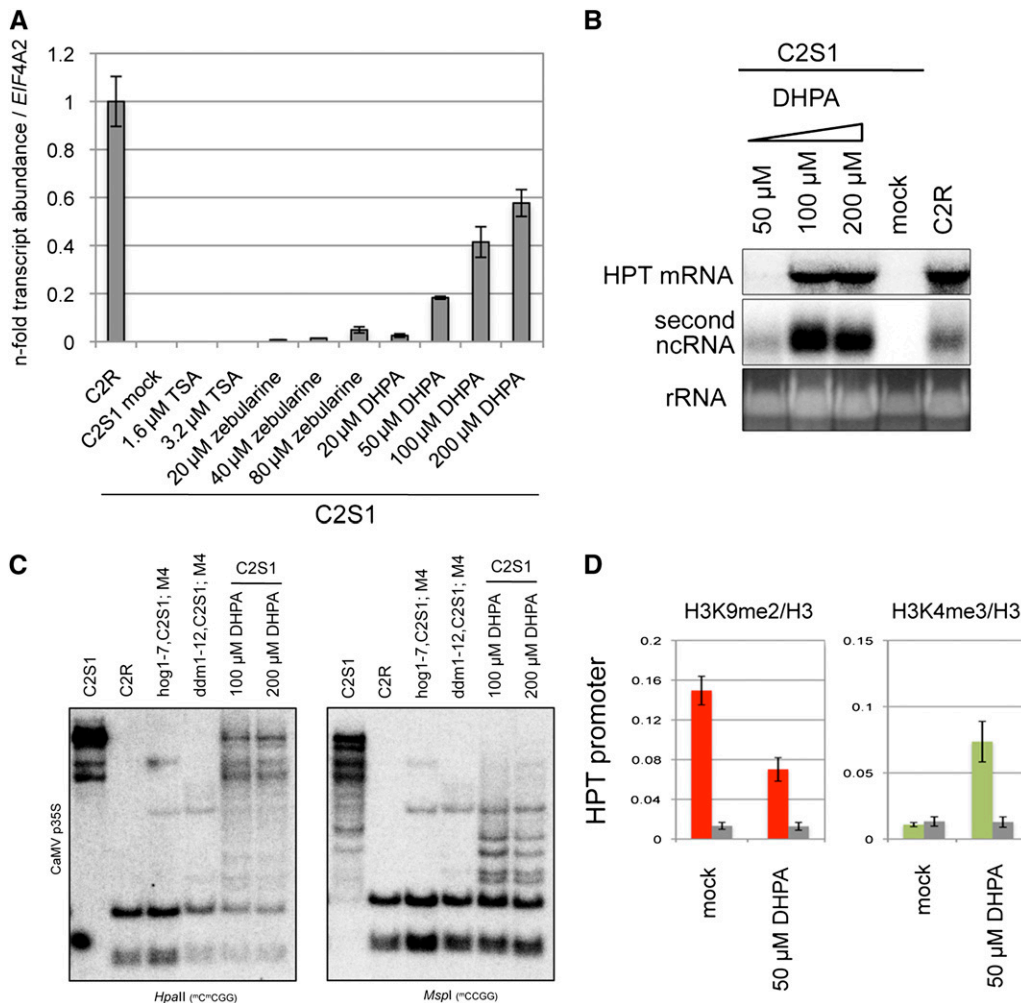
**Figure 6.** The SAHH Inhibitor DHPA Interferes with Maintenance of paTGS at the HPT Transgene.

**(A)** and **(B)** *HPT* transcript abundance in the inactive line C2S1 is significantly increased after treatments with DHPA.
**(C)** Increasing levels of DHPA lead to hypomethylation of the P35S promoters at the silent *HPT* transgene.
**(D)** The levels of H3K9 dimethylation (red) and H3K4 trimethlyation (green) at the P35S promoter changed after SAHH inhibitor treatments. Error bars in
**(A)** and **(D)** denote SD from triplicate analyses.
[See online article for color version of this figure.]

both modifications was effective only above a certain threshold: mutant allele *ddm1-5*, isolated based on its strong reactivation of a transcriptionally silent HPT repeat (Mittelsten Scheid et al., 1998; Jeddeloh et al., 1999) but probably not a complete loss-of-function mutation (this study), did not evoke significant hygromycin resistance before the third inbred homozygous generation. Significant activity of the primary P35S promoter in the first homozygous mutant generation was only obtained with the three new *DDM1* mutations that disrupt the conserved regions of the protein and are likely more deleterious. While numerous previously mentioned studies describe the large-scale consequences of *ddm1* mutations for gene expression, transposon activation, and diverse chromatin modifications, the mechanistic connection between these effects and the remodeling activity of the protein extrapolated from in vitro experiments

(Brzeski and Jerzmanowski, 2003) still remains to be uncovered. In this context, it is interesting that we observed decreased nucleosome abundance in DDM1-deficient plants. This could link the nucleosome remodeling function of DDM1 to the maintenance of DNA and histone methylation by facilitating a permissive environment for DNA and histone methyltransferases. Since *ddm1* is frequently investigated in the context of histone modifications (Gendrel et al., 2002; Lippman et al., 2004; Habu et al., 2006), lower nucleosome occupancy should be considered in quantitative comparisons.

By contrast, *hog1* mutations have so far only been analyzed for their effects on specific targets (Rocha et al., 2005; Mull et al., 2006) and general gene expression (Jordan et al., 2007). Nevertheless, the precise functional annotation of the gene product and the biochemical evidence for its role in regulating SAH levels
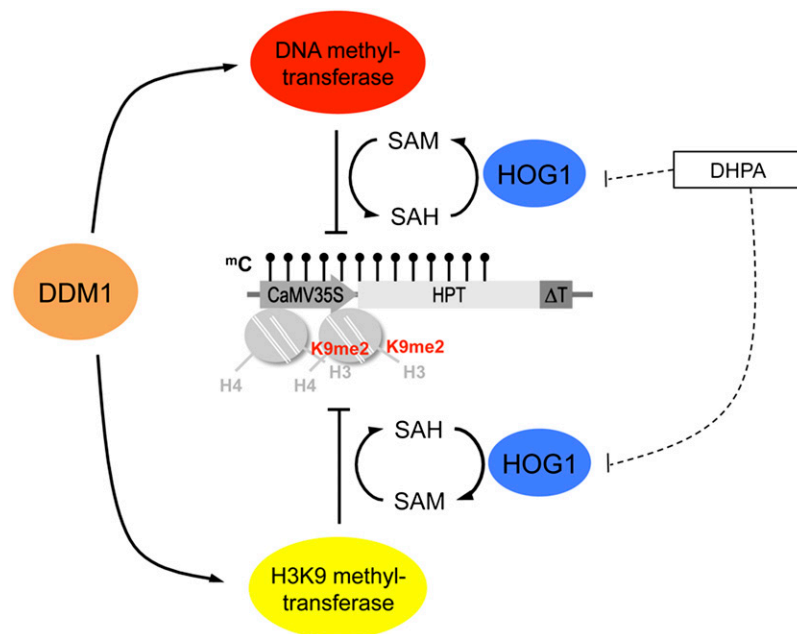
**Figure 7.** Cooperation of Multiple Chromatin Modifications to Generate Exceptional Stability of Silencing That Can Only Be Overcome by Simultaneous Removal of DNA Methylation (Black Lollipops) and Repressive Histone Modifications (Dimethylation at Lys-9 of histone H3).

The chromatin remodeling enzyme DDM1 and the *S*-adenosyl homocysteine hydrolase HOG1/SAHH are required to maintain both modifications, and only their lack in *ddm1* or *hog1* mutant or reduction of the methyl group donor SAM upon inhibitor application (DHPA) can release the tight double lock. [See online article for color version of this figure.]

(Rocha et al., 2005) make it easier to speculate about its mode of action. The methyl group donor SAM is a central hub for numerous methylation reactions modifying DNA, proteins, and metabolites (Roberts and Selker, 1995; Loenen, 2006; Roje, 2006). Therefore, substrate competition by even slightly increased SAH levels is expected to change many reactions simultaneously. The focus of HOG1 analysis has so far been on DNA as a methylation acceptor (Furner et al., 1998; Rocha et al., 2005), especially since changes in histone methylation levels were not detected in a weak *hog1* allele termed *sah*1L459F (Mull et al., 2006). Nevertheless, the new *hog1* allele brings about a substantial loss of H3K9me2 from CCs despite only slight decrease in DNA methylation, as well as a globally reduced methylation at several histones (though to different degrees). Furthermore, the mutant effects of transcriptional activation of the *HPT* transgene and endogenous transposable elements can be mimicked with a specific SAHH inhibitor. Together, these findings indicate that HOG1 is indeed a central factor in chromatin modification. This is further emphasized by the relatively small overlap of gene expression changes between *hog1-1* and treatment with the inhibitor 5-azacytidine (Jordan et al., 2007) that reduces DNA methylation and probably also 5-mdC–dependent histone methylation. Changing expression of many more genes indicates that *hog1* acts through interference with additional components. A central role of HOG1 for the plant as a whole is also evident from the severe phenotypic consequences of even subtle mutations and the embryonic lethality observed in loss-of-function mutants (Rocha et al., 2005). Due to the central role of SAM, there are

probably many more, non-chromatin-related factors involved. It should be emphasized that the SAM:SAH ratio may also be modified by metabolic regulation or by sulfur availability (Nikiforova et al., 2006). Thus, the dependence of several chromatin components on the levels of SAM and SAH offers a path by which environmental or nutritional cues can inscribe a signature in the epigenetic outfit of the genome.

The forward genetic screen for reactivation of the *HPT* allele resulted in only 21 primary mutant candidates, surprisingly few for a mutant population derived from 20,000 independent T-DNA transformation events compared with other screens following insertional mutagenesis (Budziszewski et al., 2001). In addition, several candidates turned out to carry mutations within the marker gene itself (A. Foerster, unpublished data). This, together with finding three alleles of the *DDM1* gene, indicates saturation of the screen for *trans*-acting mutations. The mechanism of epigenetic control depicted in the double lock model makes these results nevertheless plausible: the need to eliminate two different chromatin modifications simultaneously requires either rare double mutations in two independent pathways or single mutations affecting the two modifications equally, making the screen a very stringent quest for strong modifications. Although very different in their assumed mode of action, DDM1 and HOG1 fulfill the latter conditions. Therefore, the double lock model is not only supported by the molecular data, but also by the general outcome of the forward screen.

It could be asked whether the data presented here, based mainly but not exclusively on the *HPT* transgene, have relevance

beyond this particular situation. The advantage of this experimental system is that it represents a gene whose activity is absolutely nonessential for the plants unless under selection and thereby does not bias the propagation or segregation of either the active or inactive states. It is inserted in an intergenic region (Mittelsten Scheid et al., 2003) and is therefore unlikely to cause an insertional mutation. The random rearrangement producing a duplication of the P35S promoter during the initial transformation event even allowed these two regulatory elements to be compared, with the surprising result that the identical sequences, in the same genomic location and with a distance of only 2 kb between them, respond quite differently to mutations and inhibitor effects. As pointed out before (Rocha et al., 2005), the silencing system in plants was not invented to inactivate man-made transgenes. Along this line, we demonstrated a clear epigenetic effect of the SAHH inhibitors and the *hog1* and *ddm1* mutant alleles on individual endogenous targets. A significant overlap of genes differentially regulated in both mutants, mainly but not exclusively transposons (T. Baubec and O. Mittelsten Scheid, unpublished data) further indicates more sequences under double control and a significant relevance of tight silencing beyond the *HPT* transgene.

More important is thinking about the role of polyploidy in generating a stable epiallele. While a diploid progenitor line containing the very same transgene always maintained high expression, partial or complete silencing was found in several independent autotetraploid derivatives (Mittelsten Scheid et al., 2003). However, these were generated by protoplast culture and regeneration, leaving other parameters, such as hormone effects, tissue culture conditions, or even propagation of preexisting epigenetic states in individual cells, as possible sources of silencing, rather than polyploidization. Nevertheless, an association with polyploidy is very likely based on the *trans*-acting silencing between inactive and active epialleles, which is limited to tetraploid hybrids (Mittelsten Scheid et al., 2003), and with a specific set of genes that are differentially expressed in the tetraploid lines. Although gene expression changes in autotetraploids are less frequent compared with freshly formed allo-polyploids (Wang et al., 2004), polyploidization is recognized as being a significant source of genetic as well as epigenetic changes in many different plant species (for review, see Osborn et al., 2003; Adams and Wendel, 2005). paTGS can apparently generate very tightly controlled epialleles with an extremely low frequency of reversion and with the potential to be propagated and even spread among plant populations. It should be considered to be an important source of epigenetic diversity with an evolutionary impact.

## METHODS

### Plant Growth and Chemical Treatments

Stratified seeds were surface-sterilized with 5% sodium hypochlorite and 0.05% Tween 80 for 6 min and washed and air-dried overnight. Sterilized seeds were germinated and grown in Petri dishes containing agar-solidified germination medium in growth chambers under 16-h-light/8-h-dark cycles at 21°C. For treatments with hygromycin (Calbiochem), TSA (Sigma-Aldrich), ZEB (Sigma-Aldrich), and DHPA (donated by Ales

Kovarik), seeds were sown and grown directly on drug-containing plates under the conditions described above. Hygromycin (10 μg/mL), zebularine (20, 40, and 80 μM), and DHPA (50, 100, and 200 μM) in aqueous solution or TSA (1.6 and 3.2 μM) dissolved in DMSO were added to the germination medium before solidifying.

### Mutant Screen and Mapping

Diploid C2S1 plants (in the background of accession Zürich) were mutagenized by random T-DNA insertion after *Agrobacterium tumefaciens* transformation with p1'barbi (Mengiste et al., 1997). M2 seeds from 20,000 mutant M1 plants were harvested in pools of 15 M1 plants and selected on hygromycin-containing medium. *HPT*-expressing and non-expressing lines, C2R and C2S1, were used as positive and negative controls, respectively. Hygromycin-resistant plants were further propagated, and hygromycin resistance was followed in subsequent generations after selfing and outcrossing to the wild type. Sequences flanking the T-DNA insertion that were genetically linked with the mutations (*ddm1-13* and *hog1-7*) were identified by thermal asymmetric interlaced PCR as described (Liu et al., 1995). Other mutations were identified by sequencing of candidate genes (as in the case of *ddm1-11* and *ddm1-12*).

### Nucleic Acid Isolation and Gel Blot Analysis

Pools of 50 to 100 seedlings (age as indicated for the individual experiments) were shock-frozen in liquid nitrogen, homogenized, and subsequently used for DNA or RNA extraction using Phytopure (Amersham) or RNAeasy (Qiagen) kits, respectively.

For DNA methylation analysis by DNA gel blot, 10 μg of genomic DNA were digested overnight with 1 to 2 units of *Hpa*II (sensitive to $^mC^mCGG$) or *Msp*I (sensitive to $^mCCGG$) restriction enzymes. Subsequently, samples were electrophoretically separated on TAE agarose gels, depurinated for 10 min in 250 mM HCl, denatured for 30 min in 0.5 M NaOH and 1.5 M NaCl, and neutralized twice in 0.5 M Tris, 1.5 M NaCl, and 1 mM EDTA at pH 7.2 for 15 min. For RNA gel blot analysis, 10 μg of total RNA were denatured with 15% glyoxal and DMSO for 1 h at 50°C and separated on 1.4% agarose gels in 10 mM sodium phosphate buffer, pH 7.0, in a Sea2000 circular flow electrophoresis chamber (Elchrom Scientific). DNA and RNA gels were blotted onto Hybond N$^+$ membranes (GE Healthcare) overnight with 20× SSC and washed, and the samples were UV cross-linked using a Stratalinker (Stratagene). Hybridization was performed as described (Church and Gilbert, 1984). Radioactively labeled sequence-specific probes were synthesized from 25 ng of template DNA in the presence of 50 μCi of [α-$^{32}$P]dCT- (Hartmann Analytic) using the Rediprime labeling kit (Amersham). Signals of exposures in the linear range were detected with phosphor imager screens (Bio-Rad) and scanned with a Molecular Imager FX (Bio-Rad).

### Quantification of Global DNA Methylation

Total cytosine methylation was determined by cation exchange HPLC as described by Rozhon et al. (2008). All samples were analyzed in triplicate, and 5-mdC values were expressed as a percentage of total cytosine.

### Reverse Transcription and Quantitative Real-Time PCR

RNA samples were treated with 5 units of DNase I (MBI Fermentas), 0.4 units of RNasin, and 4 μL of 10× DNase I buffer for 40 min at 37°C to remove residual DNA contamination, extracted with phenol:chloroform (24:1), and subsequently ethanol-precipitated. Reverse transcription was performed on 1 μg of RNA with 0.2 μg of random hexamer primers (MBI Fermentas) using 1 unit of RevertAid H Minus M-MuLV-RTase (MBI Fermentas) at 42°C for 11/2 h. Real-time PCR analysis was performed with the 2× SensiMix Plus SYBR and Fluorescein Kit (Quantace) protocol

using an iQ5 real-time PCR system (Bio-Rad Laboratories). Ct values were analyzed using Excel (Microsoft). The primer sequences are listed in Supplemental Table 1 online.

### In Situ GUS Detection

GUS activity was detected by staining plant tissue in 0.1 M sodium phosphate buffe, pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 100 μg/mL chloramphenicol, 2 mM potassium ferrocyanide, 2 mM potassium ferri-cyanide, and 0.5 mg/mL X-glucuronide after 30 min vacuum infiltration and overnight incubation at 37°C. Subsequent washes with 70% ethanol at 37°C were performed to remove chlorophyll and enhance contrast. All samples were analyzed using a Leica MZ16FA binocular microscope with a Leica DFC300FX CCD camera. Images were acquired with the Leica Application Suite and processed with Adobe Photoshop (Adobe).

### Immunofluorescence

For the preparation of nuclei, 21-d-old plantlets were rinsed in 10 mM Tris buffer, pH 7.5, fixed by vacuum infiltration in 4% formaldehyde/Tris buffer, rinsed in Tris buffer, chopped in 500 μL chromosome isolation buffer (15 mM Tris, 2 mM Na$_2$EDTA, 0.5 mM spermin, 80 mM KCl, 20 mM NaCl, 15 mM β-mercaptoethanol, and 0.1% Triton X-100, pH 7.5), and filtered through a 50-μm nylon mesh. Fifty microliters of suspension was transferred onto microscope slides, and nuclei were attached to the slide using a cytospin centrifuge (MPW) at 2500 rpm for 10 min.

Immunolocalization of methylated cytosine was performed as described (Jasencakova et al., 2000), with minor modifications. In brief, slides were treated with pepsin (50 μg/mL in 0.01 M HCl; Roche) at 38°C (1 to 2 min), postfixed in 4% formaldehyde/2× SSC, denatured in 70% formamide/2× SSC at 80°C (2 min), and cooled in ice-cold 1× PBS. After blocking (5% BSA, 0.2% Tween 20, and 4× SSC) at 37°C (30 min), the slides were incubated with primary monoclonal mouse-anti-5-methylcytosine (Euro-gentec) and secondary goat-anti-mouse-Alexa488 (Molecular Probes) antibodies. Immunolocalization of histone H3 modifications was performed as previously described (Jasencakova et al., 2000). Slides were fixed in 4% paraformaldehyde/PBS for 20 min and blocked (5% BSA, 0.2% Tween 20, and 4× SSC) at 37°C (30 min). The slides were incubated overnight at 4°C with primary antibodies specific to H3K9me2 (T. Jenuwein; 4677) and secondary goat-anti-rabbit-AF488 (Molecular Probes). The slides were counterstained with 4',6-diamidino-2-phenylindole (Vector Laboratories) and analyzed using a Zeiss Axioplan 2 epifluorescence microscope. Monochromatic images were acquired with MetaVue (Universal Imaging) and processed with Adobe Photoshop (Adobe).

### Bisulfite Conversion, Sequencing, and Evaluation

After treatment with RNase A and proteinase K, 1 to 2 μg of genomic DNA were digested overnight with BamHI (MBI Fermentas). Subsequent bisulfite conversion was performed using the Epitect conversion kit (Qiagen) and controlled for completion as described (Hetzl et al., 2007). Converted DNA was used for PCR amplification (see Supplemental Table 1 online). PCR-amplified DNA was cloned using CloneJet (MBI Fermentas) and ligation mixes transformed into DH5α cells (Invitrogen), sequenced by terminal labeling using BigDye Terminator v3.1 (Applied Biosystems), and read at vbcgenomics.com. The sequence information obtained was analyzed with CyMATE (www.gmi.oeaw.ac.at/cymate; Hetzl et al., 2007) and Excel (Microsoft).

### ChIP

ChIP was performed as described (www.epigenome-noe.net/researchtools/protocol.php) using 3-week-old seedlings. The chromatin was immunoprecipitated with antibodies to histone H3 (Abcam; ab1791), H3K4me3 (Upstate; 07-473), H4pentaAc (Millipore; P62805), H3K9me2

(T. Jenuwein, 4677; Abcam, ab1220). Immunoprecipitated DNA was purified using a Qiagen PCR purification kit and eluted in 50 μL of EB buffer. Quantitative real-time PCR was performed in a total reaction volume of 25 μL, and quantitative PCR conditions were according to the 2× SensiMix Plus SYBR and Fluorescein kit (Quantace) protocol using an iQ5 real-time-PCR system (Bio-Rad Laboratories). Quantitative PCR data were evaluated as a ratio either to input DNA or to H3 abundance (Haring et al., 2007), as indicated.

### Immunoblot Analysis

Immunoblot analysis was performed as described (Yan et al., 2007). Approximately 20 μg per sample were loaded on 15% SDS-PAGE gels and subsequently blotted onto polyvinylidene fluoride membranes (Amersham). The primary antibodies were H3 (Abcam; ab1791), H3K9me2 (T. Jenuwein; 4677), and H3K4me3 (Upstate; 07-473); the secondary antibody was peroxidase-conjugated goat-anti-rabbit (Jackson Immuno Research). Detection was performed using Lumi-Light protein gel blotting substrate (Roche).

### Accession Numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: *DDM1*, At5g66750; *SAHH1*, At4g13940; *SAHH2*, At3g23810; *LINE1-4*, At2g01840; *EIF4A2*, At1g54270; and *TUBULIN8*, At5g23860.

### Author Contributions

The experiments were designed by T.B. and O.M.S. and performed by T.B. (main), A.P. (cytology), B.R. (protein gel blot), W.R. (global cytosine methylation), and B.W. (mutant screen). The data were analyzed by T.B., H.Q.D., A.v.H., and O.M.S. The article was written by T.B. and O.M.S.

### Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure 1.** *HPT* Transcript Abundance in Zebularine- and Mock-Treated Wild-Type and Mutant Plants.

**Supplemental Figure 2.** Location and Organization of the *HPT* Transgenic Insert in the Short Arm of Chromosome III.

**Supplemental Figure 3.** Effects of *ddm1-12* and *hog1-7* on *LINE1-4* Silencing and Global Histone Modifications.

**Supplemental Figure 4.** Treatments with the SAHH Inhibitor DHPA Releases Silencing of Repetitive Transgenes.

**Supplemental Figure 5.** Treatment with the SAHH Inhibitor DHPA Reduces Levels of DNA and Histone Methylation at Endogenous Repeats.

**Supplemental Figure 6.** DHPA Treatment Reduces H3K9me2 but Not DNA Methylation at Chromocenters.

**Supplemental Table 1.** DNA Sequence of Primers Used throughout This Study.

## REFERENCES

**Adams, K.L., and Wendel, J.F.** (2005). Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. **8:** 135–141.

**Baubec, T.** (2008). Maintenance of polyploidy-associated transcriptional gene silencing in *Arabidopsis thaliana*, dissertation (Austria: University of Vienna). http://othes.univie.ac.at/3756/.

**Baubec, T., Pecinka, A., Rozhon, W., and Mittelsten Scheid, O.** (2009). Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. Plant J. **57:** 542–554.

**Bork, P., and Koonin, E.V.** (1993). An expanding family of helicases within the 'DEAD/H' superfamily. Nucleic Acids Res. **21:** 751–752.

**Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J.** (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science **322:** 1387–1392.

**Brzeski, J., and Jerzmanowski, A.** (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. J. Biol. Chem. **278:** 823–828.

**Budziszewski, G.J., et al.** (2001). Arabidopsis genes essential for seedling viability: Isolation of Insertional mutants and molecular cloning. Genetics **159:** 1765–1778.

**Chan, S.W.L., Henderson, I.R., Zhang, X.Y., Shah, G., Chien, J.S.C., and Jacobsen, S.E.** (2006). RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in Arabidopsis. PLoS Genet. **2:** 791–797.

**Chandler, V.L., Eggleston, W.B., and Dorweiler, J.E.** (2000). Paramutation in maize. Plant Mol. Biol. **43:** 121–145.

**Chang, S., and Pikaard, C.** (2005). Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation. J. Biol. Chem. **280:** 796–804.

**Chen, Z.J., and Pikaard, C.S.** (1997). Epigenetic silencing of RNA polymerase I transcription: A role for DNA methylation and histone modification in nucleolar dominance. Genes Dev. **11:** 2124–2136.

**Chen, Z.J., and Tian, L.** (2007). Roles of dynamic and reversible histone acetylation in plant development and polyploidy. Biochim. Biophys. Acta **1769:** 295–307.

**Chinnusamy, V., and Zhu, J.K.** (2009). Epigenetic regulation of stress responses in plants. Curr. Opin. Plant Biol. **12:** 133–139.

**Chong, S.Y., and Whitelaw, E.** (2004). Epigenetic germline inheritance. Curr. Opin. Genet. Dev. **14:** 692–696.

**Church, G.M., and Gilbert, W.** (1984). Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995.

**Cubas, P., Vincent, C., and Coen, E.** (1999). An epigenetic mutation responsible for natural variation in floral symmetry. Nature **401:** 157–161.

**Elmayan, T., Proux, F., and Vaucheret, H.** (2005). Arabidopsis RPA2: A genetic link among transcriptional gene silencing, DNA repair, and DNA replication. Curr. Biol. **15:** 1919–1925.

**Finnegan, E.J.** (2002). Epialleles - A source of random variation in times of stress. Curr. Opin. Plant Biol. **5:** 101–106.

**Foerster, A.M.** (2009). Functional analysis of epialleles in diploid and tetraploid *Arabidopsis thaliana*, dissertation (Austria: University of Vienna). http://othes.univie.ac.at/7879/1/2009-10-27_0001330.pdf.

**Fuchs, J., Demidov, D., Houben, A., and Schubert, I.** (2006). Chromosomal histone modification patterns: From conservation to diversity. Trends Plant Sci. **11:** 199–208.

**Furner, I.J., Sheikh, M.A., and Collett, C.E.** (1998). Gene silencing and homology-dependent gene silencing in Arabidopsis: Genetic modifiers and DNA methylation. Genetics **149:** 651–662.

**Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A.** (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science **297:** 1871–1873.

**Habu, Y., Mathieu, O., Tariq, M., Probst, A.V., Smathajitt, C., Zhu, T., and Paszkowski, J.** (2006). Epigenetic regulation of transcription in intermediate heterochromatin. EMBO Rep. **7:** 1279–1284.

**Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C., and Stam, M.** (2007). Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. Plant Methods **3:** 11.

**Hetzl, J., Foerster, A.M., Raidl, G., and Mittelsten Scheid, O.** (2007). CyMATE: A new tool for methylation analysis of plant genornic DNA after bisulphite sequencing. Plant J. **51:** 526–536.

**Jasencakova, Z., Meister, A., Walter, J., Turner, B.M., and Schubert, I.** (2000). Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. Plant Cell **12:** 2087–2100.

**Jeddeloh, J.A., Stokes, T.L., and Richards, E.J.** (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat. Genet. **22:** 94–97.

**Johnson, L.M., Cao, X.F., and Jacobsen, S.E.** (2002). Interplay between two epigenetic marks: DNA methylation and histone H3 lysine 9 methylation. Curr. Biol. **12:** 1360–1367.

**Johnson, L.M., Law, J.A., Khattar, A., Henderson, I.R., and Jacobsen, S.E.** (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet. **4:** e1000280.

**Jordan, N.D., West, J.P., Bottley, A., Sheikh, M., and Furner, I.** (2007). Transcript profiling of the hypomethylated hog1 mutant of Arabidopsis. Plant Mol. Biol. **65:** 571–586.

**Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K., and Richards, E.J.** (1996). Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proc. Natl. Acad. Sci. USA **93:** 12406–12411.

**Kakutani, T., Munakata, K., Richards, E.J., and Hirochika, H.** (1999). Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of *Arabidopsis thaliana*. Genetics **151:** 831–838.

**Kalisz, S., and Purugganan, M.D.** (2004). Epialleles via DNA methylation: Consequences for plant evolution. Trends Ecol. Evol. **19:** 309–314.

**Kato, M., Miura, A., Bender, J., Jacobsen, S.E., and Kakutani, T.** (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr. Biol. **13:** 421–426.

**Kovarik, A., Koukalova, B., Holy, A., and Bezdek, M.** (1994). Sequence-specific hypomethylation of the tobacco genome induced with dihydroxypropyladenine, ethionine and 5-azacytidine. FEBS Lett. **353:** 309–311.

**Kovarik, A., Koukalova, B., Lim, K.Y., Matyasek, R., Lichtenstein, C.P., Leitch, A.R., and Bezdek, M.** (2000b). Comparative analysis of DNA methylation in tobacco heterochromatic sequences. Chromosome Res. **8:** 527–541.

**Kovarik, A., Van Houdt, H., Holy, A., and Depicker, A.** (2000a). Drug-induced hypomethylation of a posttranscriptionally silenced transgene locus of tobacco leads to partial release of silencing. FEBS Lett. **467:** 47–51.

**Lippman, Z., et al.** (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430:** 471–476.

**Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R.** (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol. **1:** 420–428.

**Liu, Y.G., Mitsukawa, N., Oosumi, T., and Whittier, R.F.** (1995). Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J. **8:** 457–463.

**Loenen, W.A.M.** (2006). S-adenosylmethionine: Jack of all trades and master of everything? Biochem. Soc. Trans. **34:** 330–333.

**Madlung, A., and Comai, L.** (2004). The effect of stress on genome regulation and structure. Ann. Bot. (Lond.) **94:** 481–495.

**Martinez-Zapater, J.M., Estelle, M.A., and Somerville, C.R.** (1986). A highly repeated DNA-sequence in *Arabidopsis thaliana*. Mol. Gen. Genet. **204:** 417–423.

**Masterson, J.** (1994). Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. Science **264:** 421–424.

**Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.V., Colot, V., Schubert, I., and Tourmente, S.** (2003). Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in Arabidopsis. Plant Cell **15:** 2929–2939.

**Mengiste, T., Amedeo, P., and Paszkowski, J.** (1997). High-efficiency transformation of Arabidopsis thaliana with a selectable marker gene regulated by the T-DNA 1' promoter. Plant J. **12:** 945–948.

**Milos, M.** (2006). A reverse genetic approach to analyse ploidy-associated gene silencing in *Arabidopsis thaliana*, dissertation (Austria: University of Vienna). http://othes.univie.ac.at/5529/.

**Mittelsten Scheid, O., Afsar, K., and Paszkowski, J.** (1998). Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. Proc. Natl. Acad. Sci. USA **95:** 632–637.

**Mittelsten Scheid, O., Afsar, K., and Paszkowski, J.** (2003). Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. Nat. Genet. **34:** 450–454.

**Morel, J.B., Mourrain, P., Beclin, C., and Vaucheret, H.** (2000). DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. Curr. Biol. **10:** 1591–1594.

**Mull, L., Ebbs, M.L., and Bender, J.** (2006). A histone methylation-dependent DNA methylation pathway is uniquely impaired by deficiency in Arabidopsis S-adenosylhomocysteine hydrolase. Genetics **174:** 1161–1171.

**Nikiforova, V.J., Bielecka, M., Gakiere, B., Krueger, S., Rinder, J., Kempa, S., Morcuende, R., Scheible, W.R., Hesse, H., and Hoefgen, R.** (2006). Effect of sulfur availability on the integrity of amino acid biosynthesis in plants. Amino Acids **30:** 173–183.

**Ooi, S.K.T., and Bestor, T.H.** (2008). The colorful history of active DNA demethylation. Cell **133:** 1145–1148.

**Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., and Martienssen, R.A.** (2003). Understanding mechanisms of novel gene expression in polyploids. Trends Genet. **19:** 141–147.

**Pfluger, J., and Wagner, D.** (2007). Histone modifications and dynamic regulation of genome accessibility in plants. Curr. Opin. Plant Biol. **10:** 645–652.

**Probst, A.V., Fransz, P.F., Paszkowski, J., and Mittelsten Scheid, O.** (2003). Two means of transcriptional reactivation within heterochromatin. Plant J. **33:** 743–749.

**Rangan, L., Vogel, C., and Srivastava, A.** (2008). Analysis of context

sequence surrounding translation initiation site from complete genome of model plants. Mol. Biotechnol. **39:** 207–213.

**Roberts, C.J., and Selker, E.U.** (1995). Mutations affecting the biosynthesis of S-adenosylmethionine cause reduction of DNA methylation in *Neurospora crassa*. Nucleic Acids Res. **23:** 4818–4826.

**Rocha, P., Sheikh, M., Melchiorre, R., Fagard, M., Boutet, S., Loach, R., Moffatt, B., Wagner, C., Vaucheret, H., and Furner, I.** (2005). The Arabidopsis HOMOLOGY-DEPENDENT GENE SILENCING1 gene codes for an *S*-adenosyl-L-homocysteine hydrolase required for DNA methylation-dependent gene silencing. Plant Cell **17:** 404–417.

**Roje, S.** (2006). S-Adenosyl-L-methionine: Beyond the universal methyl group donor. Phytochemistry **67:** 1686–1698.

**Rozhon, W., Baubec, T., Mayerhofer, J., Mittelsten Scheid, O., and Jonak, C.** (2008). Rapid quantification of global DNA methylation by isocratic cation exchange high-performance liquid chromatography. Anal. Biochem. **375:** 354–360.

**Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A.** (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136:** 461–472.

**Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F.** (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J. **21:** 6549–6559.

**Stebbins, G.L.** (1966). Chromosomal variation and evolution. Science **152:** 1463–1469.

**Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J.** (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc. Natl. Acad. Sci. USA **100:** 8823–8827.

**Vaillant, I., and Paszkowski, J.** (2007). Role of histone and DNA methylation in gene regulation. Curr. Opin. Plant Biol. **10:** 528–533.

**Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J.** (1993). *Arabidopsis thaliana* DNA methylation mutants. Science **260:** 1926–1928.

**Wang, J.L., Tian, L., Madlung, A., Lee, H.S., Chen, M., Lee, J.J., Watson, B., Kagochi, T., Comai, L., and Chen, Z.J.** (2004). Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics **167:** 1961–1973.

**Weretilnyk, E.A., Alexander, K.J., Drebenstedt, M., Snider, J.D., Summers, P.S., and Moffatt, B.A.** (2001). Maintaining methylation activities during salt stress. The involvement of adenosine kinase. Plant Physiol. **125:** 856–865.

**Xu, C.R., Liu, C., Wang, Y.L., Li, L.C., Chen, W.Q., Xu, Z.H., and Bai, S.N.** (2005). Histone acetylation affects expression of cellular patterning genes in the Arabidopsis root epidermis. Proc. Natl. Acad. Sci. USA **102:** 14469–14474.

**Yan, D., Zhang, Y., Niu, L., Yuan, Y., and Cao, X.** (2007). Identification and characterization of two closely related histone H4 arginine 3 methyltransferases in *Arabidopsis thaliana*. Biochem. J. **408:** 113–121.

**Yoshida, M., Horinouchi, S., and Beppu, T.** (1995). Trichostatin A and trapoxin: Novel chemical probes for the role of histone acetylation in chromatin structure and function. Bioessays **17:** 423–430.

**Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J., and Hornby, D.P.** (2002). Zebularine: A novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. J. Mol. Biol. **321:** 591–599.

# Publikace 12

# Epigenetic Regulation of Repetitive Elements Is Attenuated by Prolonged Heat Stress in *Arabidopsis* [W][OA]

**Ales Pecinka,[a,1] Huy Q. Dinh,[a,b] Tuncay Baubec,[a,2] Marisa Rosa,[a] Nicole Lettner,[a] and Ortrun Mittelsten Scheid[a,3]**

[a] Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, 1030 Vienna, Austria
[b] Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, University of Veterinary Medicine Vienna, 1030 Vienna, Austria

**Epigenetic factors determine responses to internal and external stimuli in eukaryotic organisms. Whether and how environmental conditions feed back to the epigenetic landscape is more a matter of suggestion than of substantiation. Plants are suitable organisms with which to address this question due to their sessile lifestyle and diversification of epigenetic regulators. We show that several repetitive elements of *Arabidopsis thaliana* that are under epigenetic regulation by transcriptional gene silencing at ambient temperatures and upon short term heat exposure become activated by prolonged heat stress. Activation can occur without loss of DNA methylation and with only minor changes to histone modifications but is accompanied by loss of nucleosomes and by heterochromatin decondensation. Whereas decondensation persists, nucleosome loading and transcriptional silencing are restored upon recovery from heat stress but are delayed in mutants with impaired chromatin assembly functions. The results provide evidence that environmental conditions can override epigenetic regulation, at least transiently, which might open a window for more permanent epigenetic changes.**

## INTRODUCTION

Terrestrial plants are inevitably exposed to temperature changes, and their sessile lifestyle requires that they deal with daily and seasonal temperature fluctuations in situ. In addition to sophisticated adaptation mechanisms for these regular variations, they have developed additional signaling, repair, and response functions that are activated upon heat stress exerted by exceptionally high temperatures. Key components of this heat response, among several other pathways involved in protecting various cellular functions and induced upon extreme heat, are heat shock proteins and their corresponding heat shock transcription factors (Kotak et al., 2007). Interestingly, heat stress leads to increased genetic instability and higher rates of somatic homologous recombination (Lebel et al., 1993; Pecinka et al., 2009). Since somatic homologous recombination is, at least partially, controlled by the configuration of chromatin at the target loci (Takeda et al., 2004; Endo et al., 2006; Kirik et al., 2006), heat stress could potentially exert its effect on genetic stability through modification of chromatin configuration and the accessibility of DNA for repair and recombination. Recently, a specific variant of histone H2A has been identified as a thermosensor, regulating temperature-dependent gene expression (Kumar and Wigge, 2010). Furthermore, it has been claimed that heat-induced acclimation can be transmitted to subsequent generations via an epigenetic mechanism (Whittle et al., 2009), although heat-induced somatic recombination rates were not elevated beyond the exposed generation (Pecinka et al., 2009). Thus, a connection between heat stress, chromatin, and epigenetically regulated gene expression is widely thought to occur but as yet has been poorly studied.

We chose to address this topic in *Arabidopsis thaliana*, which is sensitive to elevated temperatures (Binelli and Mascarenhas, 1990) and has a wide range of well-characterized epigenetic regulators and target genes (for review, see Henderson and Jacobsen, 2007). The numerous repetitive transgenic markers and endogenous repeats in *Arabidopsis* are especially suitable for studying epigenetic regulatory mechanisms. In general, expression of repeats is suppressed by transcriptional gene silencing (TGS), concomitant with high levels of DNA methylation, inactive chromatin marks, and chromatin compaction (e.g., Soppe et al., 2002). However, repetitive elements can be activated upon developmental reprogramming during pollen and seed development (Mosher et al., 2009; Slotkin et al., 2009) or due to a lack of several *trans*-acting epigenetic regulators (e.g., Lippman et al., 2003). Thus, they represent suitable indicators to score interference with epigenetic regulation under stress conditions.

Here, we show that prolonged heat stress leads to a transient transcriptional activation of transgenic as well as specific endogenous repeats that are regulated by TGS. These changes are independent of senescence, DNA repair, and heat stress signaling. Unexpectedly, heat-induced transcriptional activation does not require DNA demethylation. Whereas histone modifications

show only minor variation upon heat stress, there is evidence for a dramatic reduction in the number of nucleosomes associated with DNA. This reduction in nucleosome density is not restricted to heat stress–activated sequences but occurs throughout the genome. Efficient resilencing of some of the activated targets during a recovery phase seems to require the Chromatin Assembly Factor 1 (CAF-1) complex (Kaya et al., 2001), probably for its activity in reloading nucleosomes. Nevertheless, the higher order of heterochromatin is lost during prolonged heat stress, and this effect persists in exposed tissue beyond transcriptional resilencing.

## RESULTS

### Long Heat Stress Alleviates TGS

To investigate whether heat stress has an effect on epigenetically regulated transcription, we exposed 21-d-old in vitro grown plants of line L5, carrying a single insert of a multicopy *P35S:GUS* gene suppressed by TGS (Morel et al., 2000; Probst et al., 2004), to different regimes of elevated temperature and screened for transcriptional activation of β-glucuronidase (*GUS*) by histochemical staining. Whereas short heat stress (SHS) for 3 h at 37°C had no visible effect, very strong GUS expression was achieved with long heat stress (LHS) for 30 h at 37°C (Figure 1A). Quantitative RT-PCR revealed minor activation after SHS but more than 1000× induction after LHS (Figure 1B). The effect of LHS could not be recapitulated by multiple repetitions of SHS on subsequent days, and prior SHS did not significantly change the amount of *GUS* transcript upon subsequent LHS (Figure 1B). The same applies to *TRANSCRIPTIONALLY SILENT INFORMA-TION* (*TSI*), an endogenous family of repeats regulated by TGS (Steimer et al., 2000) (Figure 1B) and centromeric *180-bp* repeats (see Supplemental Figure 1A online). By contrast, *HEAT SHOCK PROTEIN101* (*HSP101*) was induced by a single SHS pulse and adaptively declined upon repeated SHS or LHS in all heat treatments (Figure 1C). To determine the kinetics of activation, we quantified *GUS*, *TSI*, and *HSP101* transcripts at short time intervals from 1 to 48 h at 37°C (see Supplemental Figures 1C and 1D online). As expected, *HSP101* was induced after 1 h of stress but strongly reduced at later time points despite ongoing heat treatment (see Supplemental Figure 1D online). *GUS* and *TSI* were notably activated only upon stress exposure longer than 12 or 18 h, respectively, and longer stress generally correlated with higher expression of these repeats (see Supplemental Figure 1C online). The extent and duration of activation of several marker genes were determined immediately after LHS as well as after 2 and 7 d of recovery and compared with levels in the TGS mutants *decrease in dna methylation1* (*ddm1*) (Vongs et al., 1993) and *morpheus' molecule1* (*mom1*) (Amedeo et al., 2000). LHS-induced *GUS*, *TSI*, and *180-bp* transcripts reached levels comparable to those in *mom1* but not in *ddm1* (Figure 1D; see Supplemental Figure 1B online). A recovery phase of only 2 d led to the disappearance of the majority of marker gene transcripts, revealing restoration of TGS. Therefore, TGS of several repetitive sequences can be transiently alleviated by an extended period of heat stress.

LHS was effective but permitted survival (see Supplemental Figure 1F online) and seed set. To exclude that the transcriptional activation of normally silent genes was a side effect of DNA damage and/or senescence, we assayed transcript levels of the corresponding marker genes *RAD51* (Doutriaux et al., 1998) and *OXIDOREDUCTASE* At4g10500 (Schmid et al., 2005), respectively. *RAD51* was unaffected by LHS, and the oxidoreductase was induced only after recovery when *GUS/TSI/180-bp* transcripts had already disappeared (see Supplemental Figure 1E online). Moreover, the observed activation does not depend on heat stress signaling since mutants lacking *HEAT SHOCK FACTOR A2* (*HSFA2*) (Nishizawa et al., 2006) express *TSI* and *180-bp* after LHS as efficiently as the wild type (Figure 1E). Thus, the activation of repeats is independent of DNA repair, senescence, and heat signaling.

### LHS Affects a Subset of Transcriptionally Silenced Endogenous Targets

To test the genome-wide effect on TGS targets, we performed transcriptome profiling on ATH1 Affymetrix arrays from mock- and LHS-treated plants directly (LHS R0) or after 2 d of recovery (LHS R2) and compared the results with published data from treatments for 3 h at 38°C (Kilian et al., 2007), here referred to as short heat stress (SHS R0). After LHS R0, 1058 and 1155 probe sets defining transcription units were significantly up- or downregulated (log2 fold change ≥2 or ≤−2, respectively) compared with the control. Among these, only 270 and 140 probe sets were up- or downregulated also after SHS (Figure 1F), indicating that many responses are specific for LHS. However, LHS-induced changes were transient, since only 19 (1.8%) and 9 (0.8%) genes remained up- or downregulated, respectively, after recovery. To focus on sequences that are known to be under epigenetic regulation, we extracted the data for the 1154 probe sets corresponding to repeats (Slotkin et al., 2009). These were barely affected by SHS (four each up- or downregulated) and only moderately by LHS (12 and 10 up- or downregulated) (Figure 1F, Table 1; see Supplemental Table 1 online). However, the majority (nine up- and nine downregulated) responded specifically to LHS. Reexamination of transcription of *COPIA78*, *MULE2*, *ATHI-LA6A*, *CYP40*, and *ATLANTYS2A* by quantitative RT-PCR (qRT-PCR) indeed verified significantly higher expression after LHS. With the exception of *COPIA78*, all returned to their previous levels during early recovery (Table 1; see Supplemental Figures 2A to 2C online). *COPIA78*, an long terminal repeat retrotransposon family, represents an interesting exception: it is not regulated by *DDM1* and *MOM1*, showed a strong response to LHS, and had delayed resilencing during recovery (Table 1; see Supplemental Table 1 and Supplemental Figure 2A online). *GP2NLTR*, *TA11*, *COPIA41*, and *IS112A* were downregulated by LHS and regained or even surpassed their original level of expression during recovery (Table 1; see Supplemental Figures 2E and 2F online). We further analyzed expression of *IG/LINE* and *soloLTR*, two targets of RNA-directed DNA methylation (RdDM) (Huettel et al., 2006) that are strongly activated upon mutation of *RNA-DEPENDENT RNA POLYMERASE2* (*RDR2*). After LHS, they were transcribed even more than in *rdr2* and silenced after recovery (Table 1; see Supplemental Figure 2D online). In short,
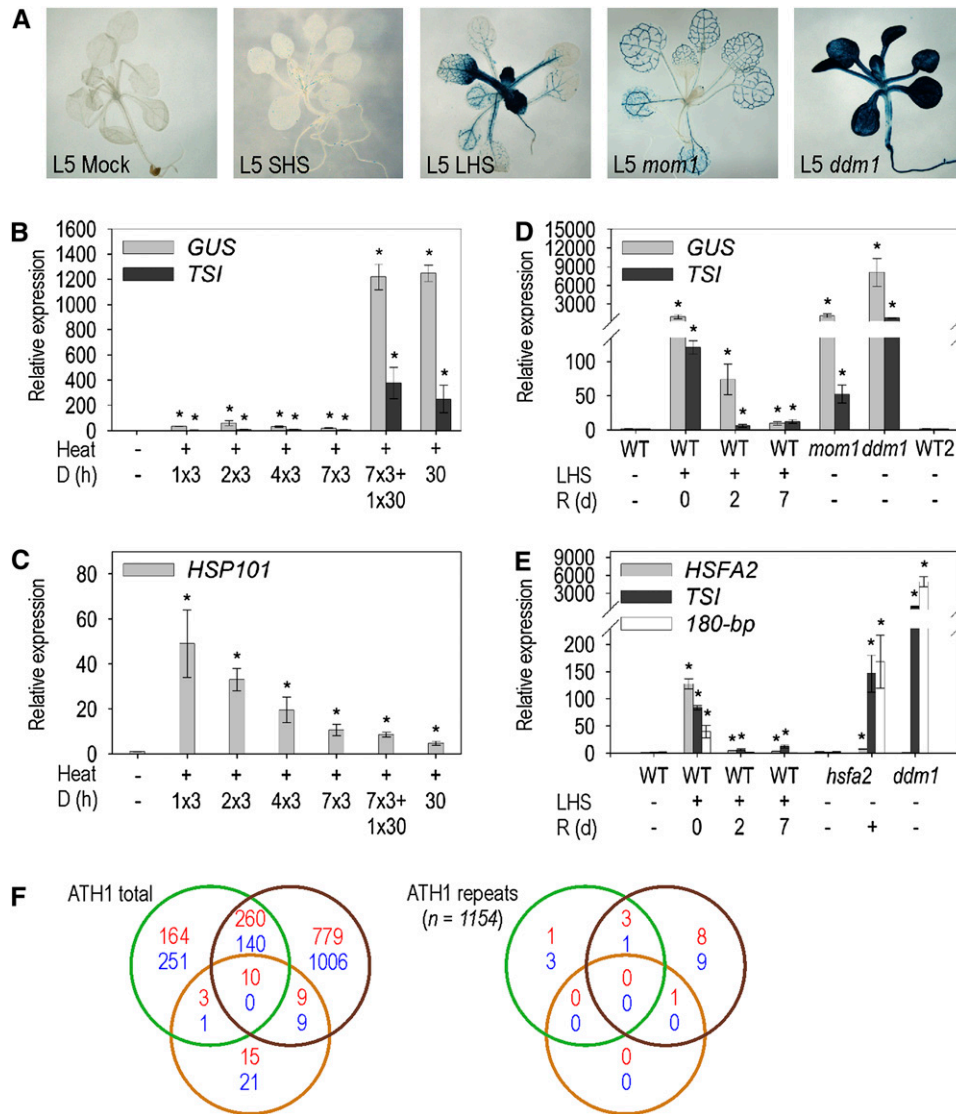
**Figure 1.** Long Heat Stress Transiently Abolishes TGS.

**(A)** GUS-stained L5 plantlets after mock, short (SHS) and long (LHS) heat stress, and nontreated after crossing to *mom1* and *ddm1* mutants.
**(B)** to **(E)** qRT-PCR for RNA of TGS targets (*GUS* and *TSI*) and heat stress marker genes (*HSP101* and *HSFA2*) in the wild type (WT = Col-0; WT2 = Col-0/Zh) and mutants (*mom1*, *ddm1*, and *hsfa2*; see text for description) after heat stress (D = frequency of application × duration in hours and R = recovery time in days), LHS = 30 h. Error bars indicate SD of triplicate measurement. Statistically significant differences between mock-treated wild types and stressed (or mutant) samples are indicated by asterisks (*t* test, P < 0.05).
**(F)** Differential gene expression (log fold changes of ≥2 [red] and ≤−2 [blue]) between mock and SHS (SHS R0, green circle) and mock versus LHS without (LHS R0, brown circle) or with (LHS R2, orange circle) recovery. ATH1 total, all probe sets; ATH1 repeats, probe sets representing repetitive elements (Slotkin et al., 2009).

LHS activated several repeats that are not transcribed after SHS. The patterns of response suggest a transient, complex, and divergent disturbance of epigenetic silencing pathways.

Another striking result from the microarray analysis was the detection of a gene cluster located close to the centromere of chromosome 2 in which 29 out of 32 genes represented on the ATH1 array were upregulated upon LHS. This cluster represents mitochondrial DNA inserted in the nuclear genome, where it has acquired some polymorphisms that allow nuclear and organelle copies to be distinguished (Stupar et al., 2001). Several mitochondrial transcripts were shown to accumulate transiently upon SHS (Adamo et al., 2008). After LHS, nuclear transcripts were also found for two of three tested genes (see Supplemental Table 2 online). The nuclear copies seem to have maintained the ancestral potential to respond to heat, but the heterochromatic neighborhood of the cluster may prevent transcription upon SHS.

**Table 1.** The Effects of LHS, *mom1*, and *ddm1* on the Transcriptional Activity of Repeats after 0, 2, and 7 d of Recovery

| Target | ORF | Transcriptional Fold Changes[a] | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Wild-Type LHS[b] | | | *mom1* | | *ddm1* | |
| | | R0 | R2 | R7 | Mock | LHS R0 | Mock | LHS R0 |
| *COPIA78* | Multiple | +++ | +++ | +++ | 0 | +++ | 0 | +++ |
| *GUS* (L5) | – | +++ | ++ | + | +++ | +++ | +++ | +++ |
| *TSI* | Multiple | +++ | + | + | +++ | +++ | +++ | +++ |
| *IG/LINE* | At5g27845 | +++ | ++ | 0 | ++ | +++ | 0 | +++ |
| *soloLTR* | – | +++ | + | 0 | +++ | +++ | 0 | +++ |
| *MULE2* | At2g15800 | +++ | 0 | 0 | +++ | +++ | +++ | +++ |
| *ATHILA* | Multiple | ++ | + | 0 | ++ | ++ | +++ | +++ |
| *HPT* (A-line) | – | ++ | 0 | 0 | + | n.d. | +++ | n.d. |
| *180-bp* | Multiple | ++ | 0 | 0 | + | ++ | +++ | +++ |
| *CYP40* | At2g15790 | + | 0 | 0 | + | + | + | + |
| *ATLANTYS2A* | At3g60930 | + | 0 | 0 | 0 | + | + | ++ |
| *IS112A* | At5g35490 | – | + | ++ | + | – | + | – |
| *COPIA4I* | At4g16870 | –[c] | 0 | - | - | – | + | ++ |
| *TA11* | At1g72920 | – | 0 | + | 0 | - | 0 | 0 |
| *GP2NLTR* | At2g15040 | – | 0 | + | 0 | - | 0 | 0 |

R, recovery; ORF, open reading frame; n.d., not determined.
[a] qRT-PCR data: +++, >400; ++, 40 to 400; +, 4 to 40; 0, −2 to 4; -, −4 to −2; −, −8 to −4.
[b] Two different wild types (WT and WT2) were included to match the different mutants as closely as possible. Unless stated otherwise, expression of the target did not differ significantly, and they are shown together.
[c] −2 to 4 for WT2.

## Transcriptional Activation Occurs Independently of DNA Demethylation

Release of TGS is often, but not obligatorily, correlated with loss of inactivating chromatin marks, such as DNA methylation and/or histone modifications. We therefore assayed both parameters after LHS. The total amount of 5-methyl deoxycytidine, reduced to one-third in *ddm1*, was at wild-type levels (6.4%) with or without LHS (see Supplemental Figure 3A online). DNA gel blots with methylation-sensitive restriction enzymes did not reveal demethylation at *TSI*, *GUS*, or *180-bp* repeats (all highly methylated in the wild type) after LHS treatment or during recovery (Figure 2A; see Supplemental Figure 3B online). Even the CG-containing transcription factor binding site in the cauliflower mosaic virus 35S promoter of the *GUS* gene, demethylated in *ddm1*, remains methylated despite LHS-induced transcription (see Supplemental Figure 3C online). By contrast, smaller bands indicating nonmethylated CG, CHG, and CHH sites in *COPIA78* appeared upon LHS (Figure 2A). Strikingly, maximum demethylation was reached only after 2 d of recovery when RNA levels were already declining, implying that it follows rather than precedes activation. Thus, LHS-induced activation of several TGS targets occurs despite DNA methylation, although this modification can be removed temporarily from a specific subset of targets.

## Transcriptional Activation Does Not Persist into the Next Generation

We tested whether activation of the TGS markers in heat-exposed plants would also affect their progeny. However, no transcriptional activation was detected for *TSI*, *GUS*, or *COPIA78* in the first poststress generation (S1) of mock- and LHS-treated plants (see Supplemental Figure 4A online). Congruently, all repeats were fully methylated in S1, including the originally demethylated *COPIA78* (see Supplemental Figure 4B online). This suggests that heat stress–induced transcriptional activation is not heritable, even for the exceptional sequences that had partially lost DNA demethylation upon stress treatment.

## Heat Stress Reduces Nucleosome Occupancy

We analyzed the chromatin of LHS-treated plants by chromatin immunoprecipitation (ChIP) for hallmarks of inactive repeats, the presence of lysine 9 dimethylation (H3K9me2), and lack of lysine 4 trimethylation (H3K4me3) at histone H3 subunits (Fuchs et al., 2006). As described (Gendrel et al., 2002), histones at repeats in *ddm1* lose H3K9me2 and gain H3K4me3; this includes the promoters of the nonactivated *COPIA78* and *HSP101* (Figure 2B; see Supplemental Figure 5 online). A significant reduction of H3K9me2, but no gain of H3K4me3, was observed directly after LHS (Figure 2B; see Supplemental Figure 5 online). Remarkably, ChIP with antibodies recognizing H3 irrespective of modifications revealed reduced nucleosome loading in *ddm1*, but also after LHS. All the *Arabidopsis* sequences analyzed had partially lost H3 association, regardless of whether they were transcribed, remained silent, or were intergenic (Figure 2B; see Supplemental Figure 5B online). An independent experiment using an antibody recognizing histone H4 (see Supplemental Figure 6 online) gave a similar result, indicating that the loss was not specific for H3 but rather was due to reduced overall nucleosome occupancy. The loss of nucleosomes was transient; all analyzed target sequences regained H3 and H4 association fully or to a large
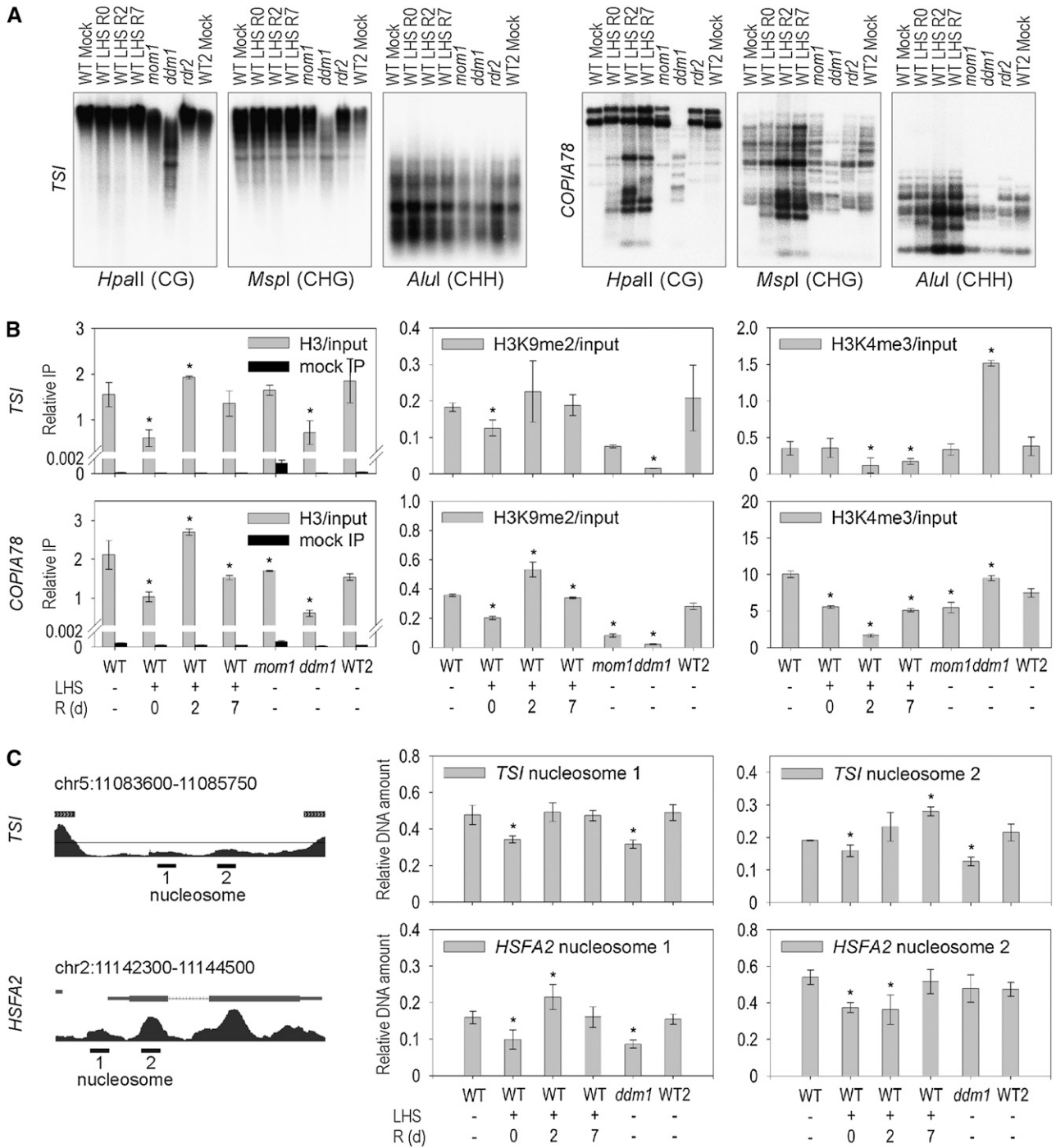
**Figure 2.** Chromatin Analysis after LHS.

**(A)** Methylation analysis of *TSI* and *COPIA78* by DNA gel blotting of LHS samples without (LHS R0) or with recovery for 2 or 7 d (LHS R2 and LHS R7).

**(B)** Histone H3 occupancy and modifications (H3K9me2 and H3K4me3), relative to input, were assessed by ChIP and qPCR.

**(C)** Nucleosome occupancy analysis by MNase I sensitivity assay at a representative TSI locus and at *HSFA2*. The positions of the PCR-amplified regions with respect to nucleosomes are indicated (left).

**(B)** and **(C)** Error bars indicate SD of triplicate measurement. R, recovery time in days. Statistically significant differences between mock-treated wild types and stressed (or mutant) samples are indicated by asterisks (*t* test, P < 0.05).

extent during recovery. In some cases, values during recovery were even higher than after mock treatments (Figure 2B). DNA fragments obtained by ChIP cover sequences in the range of 200 to 1000 bp. To obtain higher resolution, we analyzed sensitivity of defined regions by partial digestion of chromatin with Micrococcal Nuclease I (MNase I), followed by qPCR with primers located at defined nucleosome binding sites. These regions were chosen according to the genome-wide nucleosome positioning map of *Arabidopsis* (Chodavarapu et al., 2010) or on the basis of bioinformatic prediction (Segal et al., 2006). The assay confirmed reduced nucleosome occupancy at *TSI* repeats and *COPIA78* as well as at the 5′ prime regions of three genes strongly upregulated after LHS (*HSFA2*, *eEF1Balpha1*, and *UBIQUINOL-CYTO-CHROME C REDUCTASE*; Figure 2C; see Supplemental Figure 7 online). In all cases, higher sensitivity was detected immediately after LHS at the nucleosome overlapping the transcription start site and also (except for *eEF1Balpha1*) for the next nucleosome downstream (Figure 2C; see Supplemental Figures 7D and 7E online). In agreement with the ChIP data, nucleosomes tended to be reloaded, and sometimes even hyperaccumulated, during recovery (Figure 2C; see Supplemental Figures 7D and 7E online). Only an intergenic region that had reduced nucleosome occupancy, as evident from ChIP, did not show increased MNase I sensitivity. Thus, LHS causes an immediate and prevalent reduction in nucleosome occupancy, followed by reloading upon return to ambient temperatures.

## Heat Stress Causes Loss of Chromocenter Organization

The significant loss of nucleosomes after LHS prompted us to investigate global chromatin organization by fluorescence in situ hybridization (FISH). *180-bp* and *5S rDNA* repeats as well as a *HYGROMYCIN PHOSPHOTRANSFERASE* (*HPT*) multicopy transgene (all transcriptionally upregulated after LHS) form compact heterochromatic chromocenters (CCs) in >90% of interphase nuclei (Fransz et al., 2002; Probst et al., 2003). These were significantly dispersed in ~50% of nuclei from LHS-treated leaves (Figure 3). This rate is even higher than in *ddm1*, indicating substantial heterochromatin decondensation. The LHS-induced CC

dissociation was persistent throughout recovery for up to 1 week (Figure 3) when leaves started to become senescent. Interestingly, decondensation was not observed in nuclei from meristematic tissue or in leaves grown after the LHS treatment (Figure 3).

## CAF-1 Is Required for Efficient Resilencing

To identify how epigenetic regulation is reestablished after persistent heat stress, we compared the LHS response in mutants lacking well-defined epigenetic regulators. The extent of *TSI* induction by LHS and the kinetics of resilencing were similar between the wild type, *rdr2* (Figure 4A), and other RdDM mutants. Only *drd1*, which lacks a plant-specific putative chromatin remodeling factor of the SNF2 family (Kanno et al., 2004), showed enhanced LHS-induced transcription (Figure 4B). Nevertheless, the time course of resilencing in *drd1* was comparable to that in the wild type (Figure 4B), rendering involvement of RdDM unlikely. By contrast, *fas1* and *fas2* expressed LHS-induced *TSI* sequences long after these have been silenced in the wild type (Figures 4C and 4D). These mutants lack different subunits of CAF-1, which loads nucleosomes onto freshly replicated DNA (Kaya et al., 2001). Using ChIP, we tested the kinetics of nucleosome occupancy on *TSI* repeats in heat-stressed wild-type and *fas1* plants (Figure 4E). Wild-type plants lost nucleosomes immediately after stress, with the original level being restored during recovery. By contrast, *fas1* plants had already mildly reduced nucleosome occupancy in mock-treated samples. This was further reduced after LHS, and there was no recovery even after 7 d. This is in agreement with leaky *TSI* silencing in *fas1* mutants (Figures 4C and 4D) and may explain the delayed *TSI* resilencing in CAF-1 mutants, suggesting that CAF-1 is important for efficient restoration of silencing after LHS (Figure 5).

## DISCUSSION

We have shown that several classes of repetitive elements in the *Arabidopsis* genome that are silenced by epigenetic regulation at ambient temperature were transcriptionally activated upon exposure of plants to prolonged periods of heat stress. These
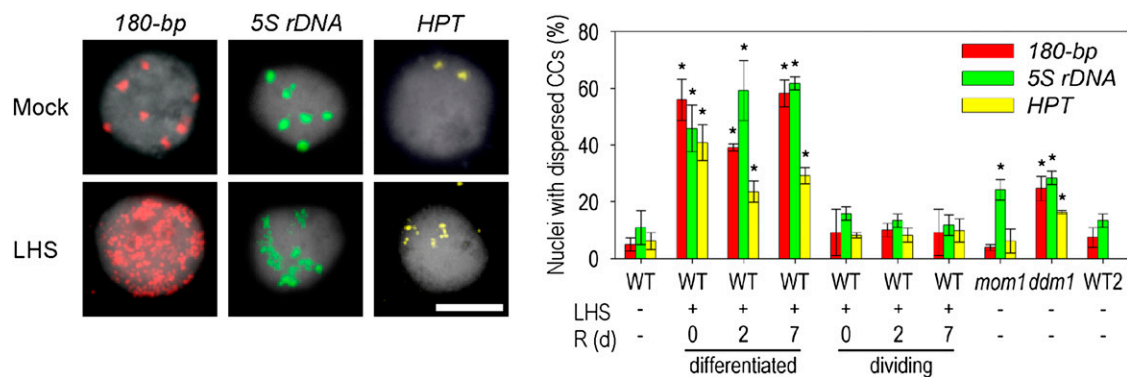


**Figure 3.** LHS Leads to Loss of Heterochromatin Compaction.

Heterochromatin condensation was analyzed by FISH with *180-bp* (red, left), *5S rDNA* (green, middle), and *HPT* (yellow, right) probes in nuclei (n = 240/experimental point) of mock- and LHS-treated plants and mutant controls. Bar = 5 μm. Error bars indicate SD of triplicate measurement. R, recovery time in days. Statistically significant differences between mock-treated wild types and stressed (or mutant) samples are indicated by asterisks (t test, P < 0.05).
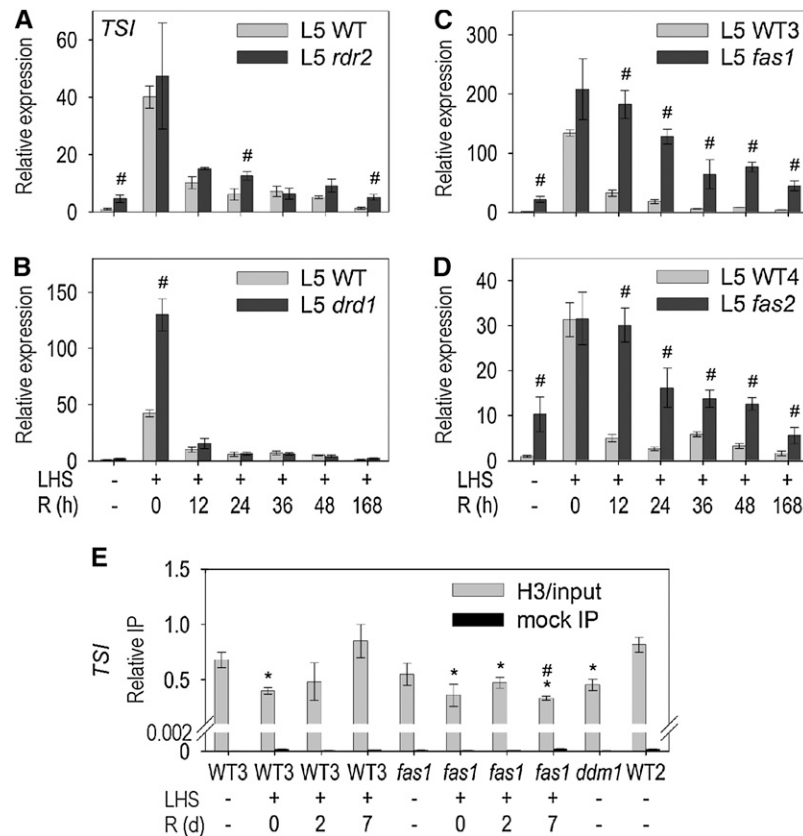
**Figure 4.** Involvement of CAF-1 in Resilencing.

**(A)** to **(D)** Kinetics of *TSI* expression after LHS quantified by qRT-PCR during recovery (R = recovery time in hours) in the wild type, RdDM, and CAF-1 mutants (see text for description). WT = Col-0, WT3 = Enk/Col-0, and WT4 = Ler/Col-0.

**(E)** Histone H3 occupancy (relative to input) was assessed by ChIP and qPCR. R = recovery time in days.

**(A)** to **(E)** Error bars indicate SD of triplicate measurement.

**(A)** to **(D)** Statistically significant differences between wild-type and mutant samples at the same time points are indicated by # (*t* test, P < 0.05).

**(E)** Statistically significant differences between mock-treated and heat-stressed plants (wild type or mutants, respectively) are indicated by asterisks (*t* test, P < 0.05).

conditions also caused differential expression of a subset of protein-coding genes. Although there was some overlap with the response to SHS pulses, the pattern and kinetics of altered expression were surprisingly different. This was not due to the detrimental effects of the prolonged application of stress, since plants could recover completely from the stress, and the transcriptional response was transient and independent of DNA damage signaling and senescence. The fact that transcriptional activation was limited to heat stress of >24 h suggests that it is a rather specific response, distinct from that of the regular diurnal changes in environmental conditions.

The consequences of long-lasting heat treatment were also distinct from genetic interference with transcriptional silencing. While several targets showed responses under heat stress similar to those of epigenetic mutants, others reacted differently. *COPIA4I* and *IS112A* elements were downregulated by long exposure to heat but were weakly affected by *mom1* and upregulated by *ddm1* (Table 1). In addition, heat stress activated the RdDM targets *IG/Line* and *soloLTR* to an extent beyond that

seen in the *rdr2* mutant. There were also unexpected differences in mechanistic aspects. In contrast with several other stress effects (reviewed in Madlung and Comai, 2004; Chinnusamy and Zhu, 2009), or upon loss of the epigenetic regulators *DDM1*, *MET1*, *HOG1*, *CMT3*, and *VIM1* (Chan et al., 2005; Woo et al., 2007), LHS-induced transcriptional activation of repeats occurred without loss of DNA methylation, thereby resembling the effect of mutations in *MOM1*, *FAS1*, *FAS2*, *BRU1*, and *RPA2* (Amedeo et al., 2000; Takeda et al., 2004; Elmayan et al., 2005). The only specifically LHS-induced element (*COPIA78*), although repetitive and with heterochromatic marks, was not expressed in the *ddm1* mutant, and demethylation here followed rather than preceded transcription. This is similar to the *Tam3* transposon of *Antirrhinum majus*, which is activated and demethylated at low temperature (15°C) and in which DNA demethylation coupled to replication is a consequence of transcriptional activation (Hashida et al., 2003, 2006). Methylation within the body of a tobacco (*Nicotiana tabacum*) gene has been shown to be reduced by heavy metal and oxidative stress (Choi and Sano,
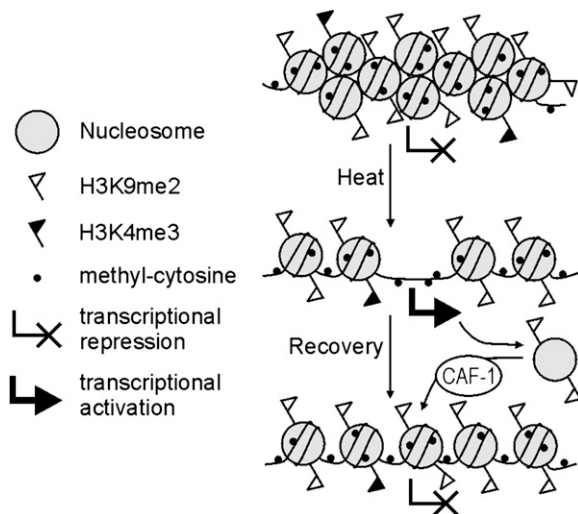
**Figure 5.** Model of Heat Stress–Induced Epigenetic Changes.

Transcriptionally inactive repeats reside in compact, heavily DNA-methylated heterochromatin with substantial H3K9 dimethylation and low levels of H3K4 trimethylation (top); after LHS, nucleosomes are partially removed rather than their modifications being altered, while heterochromatin becomes decondensed and transcriptionally active (middle). During recovery, nucleosomes are reloaded (partially via CAF-1 activity) and dimethylated at H3K9, but without reconstituting compact heterochromatin (bottom).

2007). However, it is not clear whether this is required for activation since the promoter was also unmethylated prior to stress application (Choi and Sano, 2007). Transcriptional activation without demethylation can occur also upon other stress treatments (Lang-Mladek et al., 2010). In general, neither demethylation nor removal of histone modifications appears to be essential for the activation of several repeats by heat stress. Together with the relatively unaffected (according to microarray data) expression levels of known TGS genes in LHS-treated plants, this indicates that heat stress causes a complex transcriptional response not limited to a specific pathway or factor in the regulation of repeat silencing.

Looking for common features of genes differentially expressed after LHS, it was striking that six out of 10 downregulated repeats (*COPIA4I*, *COPIA4LTR*, *IS112A*, *TA11*, *TAT1*, and *GP2NLTR*) belong to loci known to determine disease resistance, and some of these genes also had reduced transcript levels. For example, the *RECOGNITION OF PERONOSPORA PARASITICA4* (*RPP4*) locus, associated with *COPIA4* repeats, contains three assigned open reading frames: At4g16860 (*RPP4 + COPIA4LTR*), At4g16870 (*COPIA4I*), and At4g16880, all of which are downregulated after heat stress. This resembles the finding that genes within resistance clusters, including neighboring repeats, are often coregulated (Yi and Richards, 2007). Therefore, sequences in such a genomic neighborhood may be affected by LHS only indirectly and could reflect heat stress effects on the resistance genes, followed by spreading of transcriptional silencing to the close vicinity. Indeed, even moderately increased temperatures can reduce resistance to biotic stress by pathogens (Wang et al., 2009), although the expression levels of these genes were not analyzed in this study.

The upregulated and coregulated cluster of what were originally mitochondrial genes integrated into the nuclear genome may have maintained its heat response, with additional epigenetic regulation imposed by its heterochromatic environment.

Changes in histone modifications and/or expression levels of the enzymes exerting these changes have been described for stress responses in several experimental systems (reviewed in Chinnusamy and Zhu, 2009). The reduction in the inactivating chromatin mark H3K9me2 relative to the input in our experiments could be interpreted as confirmation of such a correlation, as could the small increase of H3K9me2 at some targets after 2 d of recovery, which is in agreement with an increased expression level of histone methyltransferase *KYP1* immediately after heat stress (according to microarray data). However, the quantification of H3 and H4 association and cleavage efficiency by MNase I document that prolonged heat stress resulted in a partial dissociation of histones from DNA. This would explain the apparent loss of both H3K9me2 and H3K4me3 (e.g., Figure 2B) compared with input values. Considering the reduction in nucleosome occupancy by normalizing the values to H3, it is clear that the levels of modifications on the remaining histones remained relatively unchanged. qPCR after ChIP experiments reveals nucleosome association ± 1000 bp around the primer binding sites due to the size of DNA fragments used, whereas PCR after MNase I assays reveals chromatin organization with less coverage but higher resolution. These independent assays both indicate substantial nucleosome loss at most regions analyzed. Differences between neighboring nucleosomes or remaining nucleosomes at individual targets nevertheless indicate a potential specificity of the response. Reduced nucleosome density may facilitate access of the transcriptional machinery to the promoters of repetitive elements, thus allowing their expression, similar to nucleosome depletion at *HSP70* promoters in *Drosophila melanogaster* upon heat stress (Petesch and Lis, 2008). Even more support for the role of histone-mediated transcriptional regulation in the temperature response comes from the recent discovery of the important role of the histone H2A.Z variant in *Arabidopsis* (Kumar and Wigge, 2010). At moderately high temperatures, tight wrapping of H2A.Z and the amount of H2A.Z are reduced at the promoter of heat-responsive genes, such as *HSP70*, which is associated with their increased transcriptional activity and with decreased expression of certain other targets. However, this cannot explain heat stress activation of TGS targets, since heavy DNA methylation at their promoters is mutually antagonistic with the H2A.Z modification in *Arabidopsis* (Zilberman et al., 2008). Therefore, the more extreme and lasting heat stress in our experiments seems to destabilize and/or remove entire nucleosomes, including those containing canonical histones. Our data are in agreement with the suggestion by Kumar and Wigge (2010) that the removal of nucleosomes is independent of transcription since individual nucleosomes are not removed in spite of transcription, while other, nontranscribed, parts of the genome also showed a reduction in H3 association. Whether this removal of nucleosomes is an active process or a passive response to the elevated temperature remains to be elucidated. A requirement for active reloading, in parallel to regaining epigenetic regulation of the repeats and restoring the original nucleosome loading upon recovery, is

suggested by the delayed resilencing of some repeats in mutants with reduced CAF-1 functionality, which generally have lower nucleosome density (Kirik et al., 2006).

Despite unchanged (DNA and histone methylation) or only transiently modified (transcription and histone loading) attributes, one parameter of chromatin organization was not restored to prestress conditions. The massive dissociation of heterochromatin, which exceeded even that in *ddm1* mutants (Mittelsten Scheid et al., 2002; Soppe et al., 2002), remained in nuclei of differentiated tissue that had been exposed to LHS, beyond the recovery phase when silencing and nucleosomes had been reinstalled and until exposed leaves started to show signs of senescence. Together with the general loss of nucleosomes, LHS-induced decondensation of chromocenters could increase the accessibility of DNA to transcription complexes. This seems likely in *Drosophila*, where heat stress induces puffing of chromosomes at *HSP70* loci. The process requires poly(ADP)ribose polymerase and is essential for high levels of HSP70 and thermotolerance being reached (Tulin and Spradling, 2003). Decondensed heterochromatin in *Arabidopsis* was found in 2-d-old seedlings, in response to dedifferentiation in cell culture or floral transition in development (Mathieu et al., 2003; Tessadori et al., 2007a, 2007b), but regular chromocenters were formed in a stepwise process after a longer period in culture. Heterochromatin decondensation per se was not sufficient for repeat activation (Tessadori et al., 2007b). More permanent and even repeat-specific heterochromatin decondensation has been described for plants grown at low light intensity (Tessadori et al., 2009). Since this was specific for ecotypes that originate from low geographical latitudes with naturally high light intensity, this can also be seen as a stress response. While life-long culture of these plants at higher light intensity could eliminate the phenotype of CC decondensation (Tessadori et al., 2009), the study does not address whether already decondensed chromatin could revert to the regular configuration by a switch in light conditions, which is a question of interest in the context of our data. Nevertheless, decondensation of heterochromatin is not a general response to stress, since we did not observe this phenotype after freezing ($-4°C$ for 24 h) or UV-C irradiation (3000 J/m$^2$). It also does not affect all tissues equally, since meristematic nuclei were excluded from LHS-induced decondensation. This may indicate an additional safeguarding mechanism to minimize epigenetic and possibly genetic damage in the germ line. It is possible that decondensation is a controlled process that allows increased transcriptional activity of heterochromatin-embedded targets that are important for heat stress tolerance in differentiated cells, while preventing repeat activation in dividing cells and upon the formation of subsequent generations. However, the open chromatin after prolonged heat exposure could allow occasional expression switching and may contribute to a potential influence of environmental factors on the epigenetic landscape.

## METHODS

### Plant Material and Growth Conditions

*Arabidopsis thaliana* line L5 (Morel et al., 2000; Elmayan et al., 2005) is in the Columbia-0 (Col-0) background and was crossed with *mom1-1* in Zh

(Amedeo et al., 2000), *ddm1-5* in Zh (Mittelsten Scheid et al., 1998), *fas1-1* in Enk and *fas2-1* in Landsberg *erecta* (L*er*) (Kaya et al., 2001), *rdr2-1* in Col-0 (Xie et al., 2004), and *drd1-6* in Col-0 (Kanno et al., 2004). Furthermore, we used *hsfa2-1* (Charng et al., 2007) in Col-0 and Line A (Mittelsten Scheid et al., 1998) either as wild type or crossed with *mom1-1* or *ddm1-5* (all in Zh). WT refers to Col-0. WT2/3/4 F3 hybrids between Col-0 and Zh, Enk, or L*er*, respectively, were used to match the outcrossed lines as closely as possible.

Plants were grown for 21 d after sowing on GM medium in vitro at 21°C under long-day conditions (16 h light/8 h dark) prior to stress. For heat stress, plants were transferred to 37°C for 3 h (SHS) or 30 h (LHS) starting in the light period and allowed to recover under prestress growth conditions.

### GUS Staining

GUS histochemical staining was performed as described (Pecinka et al., 2009).

### Primers

The primers used in this study are listed and their use is specified in Supplemental Table 3 online.

### DNA Methylation Analysis

Genomic DNA was isolated using a DNeasy Plant Maxi Kit (Qiagen). For DNA gel blot assays, 5 μg of DNA were digested with 20 units of *Hpa*II, *Msp*I, or *Alu*I (MBI Fermentas), separated on 1.2% agarose gels, depurinated in 250 mM HCl for 10 min, denatured in 0.5 M NaOH and 1.5 M NaCl for 30 min, and neutralized in 0.5 M Tris, 1.5 M NaCl, and 1 mM EDTA at pH 7.2 for 2× 15 min. DNA was blotted onto Hybond N$^+$ membranes (Amersham) with 20× SSC, washed, and UV cross-linked with a Stratalinker (Stratagene). Hybridization was performed as described (Church and Gilbert, 1984). Sequence-specific probes (for details, see Supplemental Table 3 online) radioactively labeled with 50 μCi of dCT-α-$^{32}$P (Amersham) were synthesized by the Rediprime II Random Prime Labeling System (GE Healthcare) and purified via G50 Probequant (Amersham) columns. Signals were detected using phosphor imager screens (Amersham) and scanned by a Molecular Imager FX (Bio-Rad). For the specific methylation assay at the ASF-1 transcription factor binding site, 200 ng of genomic DNA were digested with 5 units of *Tai*I (MBI Fermentas) and used as a template for PCR with primers qP35-TaiI-F/qP35-TaiI-R (amplicon 1), qP35-TaiI-2F/qP35-TaiI-R (amplicon 2) and qPCR-GUS-F/pPCR-GUS-R (control). Global DNA methylation quantification was performed in technical triplicate by cation-exchange HPLC as described (Rozhon et al., 2008).

### RNA Extraction and cDNA Synthesis

Total RNA was extracted using the RNeasy Plant Mini kit (Qiagen). cDNA was synthesized with random hexamer primers and the RevertAid M-MuLV Reverse Transcriptase kit (MBI Fermentas).

### qPCR Analysis

qRT-PCR analysis was performed in technical triplicate and with a minimum of two biological replicates using the SensiMix Plus SYBR kit (PEQLAB Biotechnologie) and iQ5 equipment (Bio-Rad). The expression values were calculated according to Pfaffl (2001) and normalized to the expression of the *UBC28* gene, which is not changed under heat stress conditions. For ChIP data, relative signal ratios of immunoprecipitated samples were normalized to those of corresponding input or histone H3 samples, as indicated. For MNase I sensitivity assays, the means of

individual MNase I–treated samples were multiplied by a correction factor to compensate for different amounts of DNA and compared.

## ChIP

ChIP was performed as described (http://www.epigenome-noe.net/researchtools/protocol.php?protid=13) with the antibodies rabbit polyclonal to histone H3 (Abcam; ab1791), rabbit polyclonal to histone H4 (Abcam; ab10158), mouse monoclonal to histone H3 dimethyl K9 (Abcam; ab1220), rabbit antiserum to histone H3 trimethyl K4 (Upstate; 07-473) and quantified by qPCR. Relative values were calculated with input DNA, for H4 set aside prior to immunoprecipitation (60 μL) and for H3 after mock treatment without antibody (500 μL).

## MNase I Sensitivity Assay

MNase I sensitivity assay was performed as published (Ricardi et al., 2010) with the following modifications. For chromatin isolation, 1 g of frozen tissue was homogenized to a fine powder, resuspended in 10 mL of extraction buffer 1 (0.44 M sucrose, 10 mM Tris-HCl, pH 8.0, 5 mM β-mercaptoethanol, and 1× protease inhibitor cocktail = 1 mM PMSF and 1 Complete, Mini, EDTA Free protease inhibitor tablet [Roche]/20 mL buffer), filtered through Miracloth, and centrifuged at 2880*g* for 20 min. The pellet was resuspended in 10 mL of extraction buffer 2 (0.25 M sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM MgCl$_2$, 1% Triton X-100, 5 mM β-mercaptoethanol, and 1× protease inhibitor cocktail), incubated on ice for 10 min, and centrifuged at 2100*g* for 20 min. The pellet was dissolved in 4 mL of extraction buffer 2 without Triton X-100 and centrifuged at 2100*g* for 20 min. The pellet was then dissolved in 4 mL of Percoll extraction buffer (95% v/v Percoll, 0.25 M sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM MgCl$_2$, 5 mM β-mercaptoethanol, and 1× protease inhibitor cocktail) and spun down for 10 min at 12,000*g*. The upper phase was transferred into a new tube, diluted at least five times with nuclei resuspension buffer (10% glycerol, 50 mM Tris-HCl, pH 8.0, 5 mM MgCl$_2$, 10 mM β-mercaptoethanol, and 1× protease inhibitor cocktail), and centrifuged at 12,000*g* for 10 min. The pellet was dissolved in 4 mL of nuclei resuspension buffer and centrifuged at 12,000*g* for 10 min (repeated twice).

For MNase I digestion, the pellet was dissolved in 500 μL Micrococcal nuclease buffer (50 mM Tris-HCl, pH 8.5, 5 mM Mg acetate, 25% glycerol, and 1 mM CaCl$_2$), and 100-μL aliquots were digested with 0, 3, 6, and 12 units of MNase I (Takara) at 37°C for 20 min. The reaction was terminated by adding 10 μL of 0.5 M EDTA, 20 μL of 1 M Tris-HCl, pH 6.8, and 1.5 μL of 14 mg/mL proteinase K and incubation at 45°C for 1 h.

For DNA recovery, DNA was recovered using standard phenol:chloroform extraction and precipitated with addition of yeast tRNA as a carrier. The pelleted nucleic acids were dissolved in 50 μL of water containing 10 μg/mL RNase A at 4°C overnight. Samples were analyzed by gel electrophoresis, and 10× diluted samples were used for qPCR. Quantitative analysis was performed on mock-treated samples (no nuclease, control for normalization) and samples treated with 12 units of MNase I (best preparation of mononucleosomes according to gel electrophoresis). Nucleosome-occupied regions were identified using the Methylome browser (http://epigenomics.mcdb.ucla.edu/Nuc-Seq/; Segal et al., 2006; Chodavarapu et al., 2010), and the primers were positioned within single sequencing reads.

## FISH and Microscopy

Nuclei were extracted either from whole plants or specific tissues (meristems or leaves that developed after stress treatment) as described (Pecinka et al., 2004) and transferred to slides using a Cytospin (MPW Medical Instruments). Hybridization, posthybridization washes, and FISH detection were performed as described (Pecinka et al., 2004). *180-bp* and *5S rDNA* probes were amplified and labeled with Biotin-dUTP or Digox-

igenin-dUTP via PCR using primers 180bpF/180bpR and 5SrRNAqF/5SrRNAqR, respectively. Plasmid pGL2 (Bilang et al., 1991) containing the *HPT* gene was labeled by nick translation. Microscopy was done with an AxioImager Z.1 (Zeiss), and the images were assembled in Photoshop (Adobe Systems).

## Genome-Wide Expression Profiling

Biological duplicates of total RNA samples were submitted to the microarray service of the Nottingham Arabidopsis Stock Centre (http://affymetrix.Arabidopsis.info/). The data files from hybridization to Affymetrix ATH1 microarrays were analyzed using the Bioconductor solution (www.bioconductor.org) under the R platform (www.r-project.org). The expression values were normalized by the GeneChip Robust Multiarray Averaging method (gcRMA; Wu et al., 2004). Differential gene expression analysis was performed with an empirical Bayes moderated *t* test using linear modeling (LIMMA; Smyth, 2004). The differentially expressed genes were identified by false discovery rate–corrected P values (≤0.05) and a log2 fold change cutoff (≥2, downregulated; ≤−2, upregulated). The transcriptional profiles of SHS R0 originate from previously published experiments (Kilian et al., 2007).

## Detection of Transcripts from Mitochondrial Insertion on Chromosome 2

Regions corresponding to ATH1 IDs 263504_s_at (AT2G07677+ATMG00940), 265227_s_at (AT2G07695+ATMG01280), and 257338_s_at (AT2G07711+ATMG00513) were amplified from cDNA with primers recognizing both nuclear and mitochondrial copies (see Supplemental Table 3 online). The PCR products were cloned and sequenced. Transcripts were assigned to nuclear or mitochondrial origin based on single nucleotide polymorphisms (see Supplemental Table 2 online).

## Accession Numbers

Accession numbers of sequences relevant for this article are as follows: At1g64230 (*UBC28*), At1g65470 (*FAS1*), At1g72920 (*TA11*), At1g74310 (*HSP101*), At2g07677/Atmg00940 (263504_s_at), At2g07695/Atmg01280 (265227_s_at), At2g07711/Atmg00513 (257338_s_at), At2g15040 (*GP2-NLTR*), At2g15790 (*CYP40*), At2g15800 (*MULE2*), At2g16390 (*DRD1*), At2g26150 (*HSFA2*), At3g60930 (*ATLANTYS2*), At4g05640 (*ATHILA6A*), At4g10500 (*OXIDOREDUCTASE*), At4g11130 (*RDR2*), At4g16870 (*COPIA4I*), At5g12110 (*eEF1Balpha1*), At5g20850 (*RAD51*), At5g25450 (*UBIQUINOL-CYTOCHROME C REDUCTASE*), At5g27845 (*IG/LINE*), At5g35490 (*IS112A*), and At5g64630 (*FAS2*). The microarray data are available under Gene Expression Omnibus accession number GSE18666.

## Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure 1.** LHS Transiently Abolishes TGS.

**Supplemental Figure 2.** Expression of Endogenous TGS Targets under LHS.

**Supplemental Figure 3.** DNA Methylation Analysis after LHS.

**Supplemental Figure 4.** LHS Activated TGS Targets Are Transcriptionally Silenced and DNA Is Methylated in the Next Generation.

**Supplemental Figure 5.** Analysis of Histone H3 Modification and Occupancy after LHS.

**Supplemental Figure 6.** Analysis of Histone H4 Occupancy after LHS.

**Supplemental Figure 7.** Analysis of Nucleosome Occupancy after LHS.

**Supplemental Table 1.** Repeats with Significantly Altered Expression after SHS and LHS without (R0) and after 2 d (R2) of Recovery.

**Supplemental Table 2.** Activation of Genes in a Nuclear Cluster of Mitochondrial Origin under SHS and LHS without (R0) and after 2 d (R2) of Recovery.

**Supplemental Table 3.** Primers Used in This Study.

## REFERENCES

**Adamo, A., Pinney, J.W., Kunova, A., Westhead, D.R., and Meyer, P.** (2008). Heat stress enhances the accumulation of polyadenylated mitochondrial transcripts in *Arabidopsis thaliana.* PLoS ONE **3:** e2889.

**Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O., and Paszkowski, J.** (2000). Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. Nature **405:** 203–206.

**Bilang, R., Iida, S., Peterhans, A., Potrykus, I., and Paszkowski, J.** (1991). The 3′-terminal region of the hygromycin-B-resistance gene is important for its activity in *Escherichia coli* and *Nicotiana tabacum.* Gene **100:** 247–250.

**Binelli, G., and Mascarenhas, J.** (1990). Arabidopsis: Sensitivity of growth to high temperature. Dev. Genet. **11:** 294–298.

**Chan, S.W., Henderson, I.R., and Jacobsen, S.E.** (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana.* Nat. Rev. Genet. **6:** 351–360.

**Charng, Y.Y., Liu, H.C., Liu, N.Y., Chi, W.T., Wang, C.N., Chang, S.H., and Wang, T.T.** (2007). A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis. Plant Physiol. **143:** 251–262.

**Chinnusamy, V., and Zhu, J.K.** (2009). Epigenetic regulation of stress responses in plants. Curr. Opin. Plant Biol. **12:** 133–139.

**Chodavarapu, R.K., et al.** (2010). Relationship between nucleosome positioning and DNA methylation. Nature **466:** 388–392.

**Choi, C.S., and Sano, H.** (2007). Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophospho-diesterase-like protein in tobacco plants. Mol. Genet. Genomics **277:** 589–600.

**Church, G.M., and Gilbert, W.** (1984). Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995.

**Doutriaux, M.P., Couteau, F., Bergounioux, C., and White, C.** (1998). Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana.* Mol. Gen. Genet. **257:** 283–291.

**Elmayan, T., Proux, F., and Vaucheret, H.** (2005). Arabidopsis RPA2: a genetic link among transcriptional gene silencing, DNA repair, and DNA replication. Curr. Biol. **15:** 1919–1925.

**Endo, M., Ishikawa, Y., Osakabe, K., Nakayama, S., Kaya, H., Araki, T., Shibahara, K.I., Abe, K., Ichikawa, H., Valentine, L., Hohn, B., and Toki, S.** (2006). Increased frequency of homologous recombination and T-DNA integration in Arabidopsis CAF-1 mutants. EMBO J. **25:** 5579–5590.

**Fransz, P., de Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I.** (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. Proc. Natl. Acad. Sci. USA **99:** 14584–14589.

**Fuchs, J., Demidov, D., Houben, A., and Schubert, I.** (2006). Chromosomal histone modification patterns - From conservation to diversity. Trends Plant Sci. **11:** 199–208.

**Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A.** (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science **297:** 1871–1873.

**Hashida, S., Kitamura, K., Mikami, T., and Kishima, Y.** (2003). Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in *Antirrhinum majus*. Plant Physiol. **132:** 1207–1216.

**Hashida, S.N., Uchiyama, T., Martin, C., Kishima, Y., Sano, Y., and Mikami, T.** (2006). The temperature-dependent change in methylation of the *Antirrhinum* transposon Tam3 is controlled by the activity of its transposase. Plant Cell **18:** 104–118.

**Henderson, I.R., and Jacobsen, S.E.** (2007). Epigenetic inheritance in plants. Nature **447:** 418–424.

**Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A.J.M., and Matzke, M.** (2006). Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. EMBO J. **25:** 2828–2836.

**Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M., and Matzke, A.J.M.** (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Curr. Biol. **14:** 801–805.

**Kaya, H., Shibahara, K., Taoka, K., Iwabuchi, M., Stillman, B., and Araki, T.** (2001). FASCIATA genes for chromatin assembly factor-1 in Arabidopsis maintain the cellular organization of apical meristems. Cell **104:** 131–142.

**Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K.** (2007). The AtGenExpress global stress expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J. **50:** 347–363.

**Kirik, A., Pecinka, A., Wendeler, E., and Reiss, B.** (2006). The chromatin assembly factor subunit FASCIATA1 is involved in homologous recombination in plants. Plant Cell **18:** 2431–2442.

**Kotak, S., Larkindale, J., Lee, U., von Koskull-Doring, P., Vierling, E., and Scharf, K.D.** (2007). Complexity of the heat stress response in plants. Curr. Opin. Plant Biol. **10:** 310–316.

**Kumar, S.V., and Wigge, P.A.** (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell **140:** 136–147.

**Lang-Mladek, C., Popova, O., Kiok, K., Berlinger, M., Rakic, B., Aufsatz, W., Jonak, C., Hauser, M.T., and Luschnig, C.** (2010). Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in Arabidopsis. Mol. Plant **3:** 594–602.

**Lebel, E.G., Masson, J., Bogucki, A., and Paszkowski, J.** (1993). Stress-induced intrachromosomal recombination in plant somatic cells. Proc. Natl. Acad. Sci. USA **90:** 422–426.

**Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R.** (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol. **1:** 420–428.

**Madlung, A., and Comai, L.** (2004). The effect of stress on genome regulation and structure. Ann. Bot. (Lond.) **94:** 481–495.

**Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.V., Colot, V., Schubert, I., and Tourmente, S.** (2003). Changes in 5S rDNA

chromatin organization and transcription during heterochromatin establishment in *Arabidopsis*. Plant Cell **15:** 2929–2939.

**Mittelsten Scheid, O., Afsar, K., and Paszkowski, J.** (1998). Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. Proc. Natl. Acad. Sci. USA **95:** 632–637.

**Mittelsten Scheid, O., Probst, A.V., Afsar, K., and Paszkowski, J.** (2002). Two regulatory levels of transcriptional gene silencing in Arabidopsis. Proc. Natl. Acad. Sci. USA **99:** 13659–13662.

**Morel, J.B., Mourrain, P., Beclin, C., and Vaucheret, H.** (2000). DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. Curr. Biol. **10:** 1591–1594.

**Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C.** (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature **460:** 283–286.

**Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., and Shigeoka, S.** (2006). Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. Plant J. **48:** 535–547.

**Pecinka, A., Rosa, M., Schikora, A., Berlinger, M., Hirt, H., Luschnig, C., and Mittelsten Scheid, O.** (2009). Transgenerational stress memory is not a general response in Arabidopsis. PLoS ONE **4:** e5202.

**Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J., and Schubert, I.** (2004). Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. Chromosoma **113:** 258–269.

**Petesch, S.J., and Lis, J.T.** (2008). Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. Cell **134:** 74–84.

**Pfaffl, M.W.** (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29:** e45.

**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 Mittelsten Scheid, O.** (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.

**Probst, A.V., Fransz, P.F., Paszkowski, J., and Mittelsten Scheid, O.** (2003). Two means of transcriptional reactivation within heterochromatin. Plant J. **33:** 743–749.

**Ricardi, M.M., Gonzalez, R.M., and Iusem, N.D.** (2010). Protocol: Fine-tuning of a Chromatin Immunoprecipitation (ChIP) protocol in tomato. Plant Methods **6:** 11.

**Rozhon, W., Baubec, T., Mayerhofer, J., Mittelsten Scheid, O., and Jonak, C.** (2008). Rapid quantification of global DNA methylation by isocratic cation exchange high-performance liquid chromatography. Anal. Biochem. **375:** 354–360.

**Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U.** (2005). A gene expression map of *Arabidopsis thaliana* development. Nat. Genet. **37:** 501–506.

**Segal, E., Fondufe-Mittendorf, Y., Chen, L.Y., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P.Z., and Widom, J.** (2006). A genomic code for nucleosome positioning. Nature **442:** 772–778.

**Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A.** (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136:** 461–472.

**Smyth, G.K.** (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3: Article3.

**Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F.** (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J. **21:** 6549–6559.

**Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Mittelsten Scheid, O., and Paszkowski, J.** (2000). Endogenous targets of transcriptional gene silencing in *Arabidopsis*. Plant Cell **12:** 1165–1178.

**Stupar, R.M., Lilly, J.W., Town, C.D., Cheng, Z., Kaul, S., Buell, C.R., and Jiang, J.M.** (2001). Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: Implication of potential sequencing errors caused by large-unit repeats. Proc. Natl. Acad. Sci. USA **98:** 5099–5103.

**Takeda, S., Tadele, Z., Hofmann, I., Probst, A.V., Angelis, K.J., Kaya, H., Araki, T., Mengiste, T., Mittelsten Scheid, O., Shibahara, K., Scheel, D., and Paszkowski, J.** (2004). BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis. Genes Dev. **18:** 782–793.

**Tessadori, F., Chupeau, M.C., Chupeau, Y., Knip, M., Germann, S., van Driel, R., Fransz, P., and Gaudin, V.** (2007b). Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated Arabidopsis cells. J. Cell Sci. **120:** 1200–1208.

**Tessadori, F., Schulkes, R.K., van Driel, R., and Fransz, P.** (2007a). Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis. Plant J. **50:** 848–857.

**Tessadori, F., et al.** (2009). PHYTOCHROME B and HISTONE DEACETYLASE 6 control light-induced chromatin compaction in *Arabidopsis thaliana.* PLoS Genet. **5:** e1000638.

**Tulin, A., and Spradling, A.** (2003). Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci. Science **299:** 560–562.

**Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J.** (1993). *Arabidopsis thaliana* DNA methylation mutants. Science **260:** 1926–1928.

**Wang, Y., Bao, Z.L., Zhu, Y., and Hua, J.** (2009). Analysis of temperature modulation of plant defense against biotrophic microbes. Mol. Plant Microbe Interact. **22:** 498–506.

**Whittle, C.A., Otto, S.P., Johnston, M.O., and Krochko, J.E.** (2009). Adaptive epigenetic memory of ancestral temperature regime in *Arabidopsis thaliana.* Botany **87:** 650–657.

**Woo, H.R., Pontes, O., Pikaard, C.S., and Richards, E.J.** (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. **21:** 267–277.

**Wu, Z.J., Irizarry, R.A., Gentleman, R., Martinez-Murillo, F., and Spencer, F.** (2004). A model-based background adjustment for oligonucleotide expression arrays. J. Am. Stat. Assoc. **99:** 909–917.

**Xie, Z.X., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C.** (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. **2:** 642–652.

**Yi, H., and Richards, E.J.** (2007). A cluster of disease resistance genes in *Arabidopsis* is coordinately regulated by transcriptional activation and RNA silencing. Plant Cell **19:** 2929–2939.

**Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S.** (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456:** 125–129.

## NOTE ADDED IN PROOF

Parts of this work are consistent with data described in a parallel publication.

**Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant, I.** (2010). Stress-induced activation of heterochromatic transcription. PLoS Genetics, in press.

# Publikace 13

TECHNICAL ADVANCE

# Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine

**Tuncay Baubec, Ales Pecinka, Wilfried Rozhon[†] and Ortrun Mittelsten Scheid[*]**
*Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Dr Bohr-Gasse 3, 1030 Vienna, Austria*

## Summary

**Covalent modification by methylation of cytosine residues represents an important epigenetic hallmark. While sequence analysis after bisulphite conversion allows correlative analyses with single-base resolution, functional analysis by interference with DNA methylation is less precise, due to the complexity of methylation enzymes and their targets. A cytidine analogue, 5-azacytidine, is frequently used as an inhibitor of DNA methyltransferases, but its rapid degradation in aqueous solution is problematic for culture periods of longer than a few hours. Application of zebularine, a more stable cytidine analogue with a similar mode of action that is successfully used as a methylation inhibitor in *Neurospora* and mammalian tumour cell lines, can significantly reduce DNA methylation in plants in a dose-dependent and transient manner independent of sequence context. Demethylation is connected with transcriptional reactivation and partial decondensation of heterochromatin. Zebularine represents a promising new and versatile tool for investigating the role of DNA methylation in plants with regard to transcriptional control, maintenance and formation of (hetero-) chromatin.**

**Keywords: DNA methylation, methylation inhibitor, zebularine, epigenetic regulation, transcriptional reactivation, Arabidopsis.**

## Introduction

Post-replicative modification of genomic DNA at the 5C position by methylation of cytosine residues ($^mC$) is widespread, though not universal, across a broad range of organisms. In those species that display it, DNA methylation is an important hallmark of epigenetic regulation, coupling additional, potentially heritable information to the genetic information while preserving the original DNA sequence. DNA methylation is enzymatically established by DNA methyltransferases and can cause direct transcriptional repression or an indirect effect via binding of specific proteins. In contrast to evolutionary relationships, DNA

methylation and its interpretation in mammals seems to be more similar to that found in higher plants than in any other animal class. In both groups, the level of methylated cytosines is significant, its location is specific, the group of proteins interacting with the modification is diverse and correct DNA methylation is required for regular development. Experimental interference with establishing or maintaining DNA methylation has a considerable and complex impact on vigour, morphology or gene expression, as observed with methyltransferase knockout or knockdown techniques (Finnegan *et al.*, 1996; Li *et al.*, 1992; Okano *et al.*, 1999; Ronemus *et al.*, 1996; Vongs *et al.*, 1993). Manipulation of DNA methylation has also been achieved by modification of target sequences

(Dieguez *et al.*, 1997; Klug and Rehli, 2006) or by specific inhibitors (for review see Lyko and Brown, 2005; and Yoo and Jones, 2006). While genetic modification of methylation is usually extensive and permanent, inhibitor treatments allow for partial and transient induction of methylation changes. Chemical analogues of cytosine which are incorporated into DNA are widely used inhibitors. They form covalent adducts with DNA methyltransferases, limiting their further catalytic activity (Santi *et al.*, 1983) and thereby reducing overall DNA methylation. 5-Azacytidine (5-aza) and 5-aza-2′-deoxycytidine (decitabine) are especially commonly applied inhibitors in plants and animals. Both induce hypomethylation, transcriptional reactivation and developmental effects in plant and animal systems, and have gained special attention as cancer therapeutics for malignancies that are based on erratic hypermethylation of tumour suppressor genes (for review see Christman, 2002; and Yoo and Jones, 2006). However, both drugs are extremely unstable in aqueous solution (Beisler, 1978; Constantinides *et al.*, 1977), making administration of defined doses difficult under physiological conditions. Further, both drugs have high toxicity and many side-effects (Ghoshal and Bai, 2007). The search for more stable and less toxic methylation inhibitor drugs has led to the identification of zebularine (1-(β-D-ribofuranosyl)-1,2-dihydropyrimidine-2-one; Figure 1) as a potent drug (Cheng *et al.*, 2003; Marquez *et al.*, 2005; Yoo and Jones, 2006; Yoo *et al.*, 2004; Zhou *et al.*, 2002), originally developed as a cytidine deaminase inhibitor. Acting in a similar way as 5-aza and decitabine, zebularine has a much longer half-life under physiological conditions and fewer side-effects (Cheng *et al.*, 2003). Its action in cancer models has been proven in several studies (Herranz *et al.*, 2006; Marquez *et al.*, 2005; Rao *et al.*, 2007; Scott *et al.*, 2007), although clinical trials have not yet been performed (Yoo and Jones, 2006). Given the limitations of 5-aza instability and toxicity in plant research applications as well (Weber *et al.*, 1990), and the original discovery of the demethylating and reactivating effect of zebularine in the filamentous fungus *Neurospora* (Cheng *et al.*, 2003), it is surprising that as far as we are aware no study has so far addressed the effect of zebularine on plant DNA.

We present data on DNA demethylation in the genomic DNA of *Arabidopsis thaliana* and *Medicago sativa* after application of different doses of zebularine and lengths of treatment. Furthermore, we compare the overall levels of $^mC$ as well as $^mC$ in different sequence contexts after zebularine treatment at transgenic and endogenous single-copy and repetitive sequences, and analyse the effect on transcriptional activity. The data show that zebularine is a potent dose-dependent and non-discriminative inducer of hypomethylation and transcription, and is a suitable tool for investigating the important role of DNA methylation in plants.

## Results

### Zebularine induces dose-dependent and transient growth inhibition

Since reduced DNA methylation results in abnormal plant development (Finnegan *et al.*, 1996; Jeddeloh *et al.*, 1998; Mathieu *et al.*, 2002; Ronemus *et al.*, 1996), the concentration range of potential effects of zebularine as a methylation inhibitor was established by scoring for its phenotypic effects on plant development. *Arabidopsis thaliana* (accession Zürich) was grown on media containing 0, 20, 40 and 80 μM zebularine (Figure 2a–d). Minor developmental retardation was observed 14 days after germination (dag) at a concentration as low as 20 μM zebularine (Figure 2b). The plants grew secondary roots, but were slightly delayed in growth and developed elongated true leaves when compared with mock-treated plants (Figure 2a). At 40 μM zebularine, true leaves did not expand and roots were much shorter (Figure 2c) than observed at 20 μM. At 80 μM zebularine, plants showed severe inhibition of growth; they did not develop beyond the cotyledon stage and had severely affected root growth (Figure 2d). Nevertheless, the majority of zebularine-treated plants from all concentrations could be rescued by transferring them after 14 or 21 days of treatment to inhibitor-free growth medium. Rescued plants showed complete recovery and a normal seed set. Therefore, transient exposure to zebularine concentrations up to 80 μM causes growth effects that indicate effectiveness and allow subsequent recovery of fertile plants after the treatment.

**Figure 1.** Chemical structure of cytidine, its methylated form, 5-methylcytidine and the methylation inhibitors 5-azacytidine and zebularine (adapted from Cheng *et al.*, 2003).
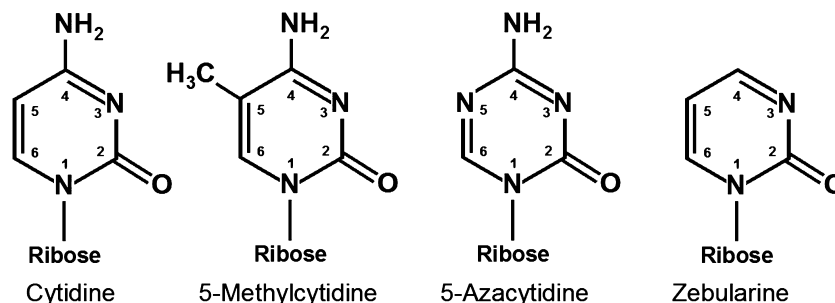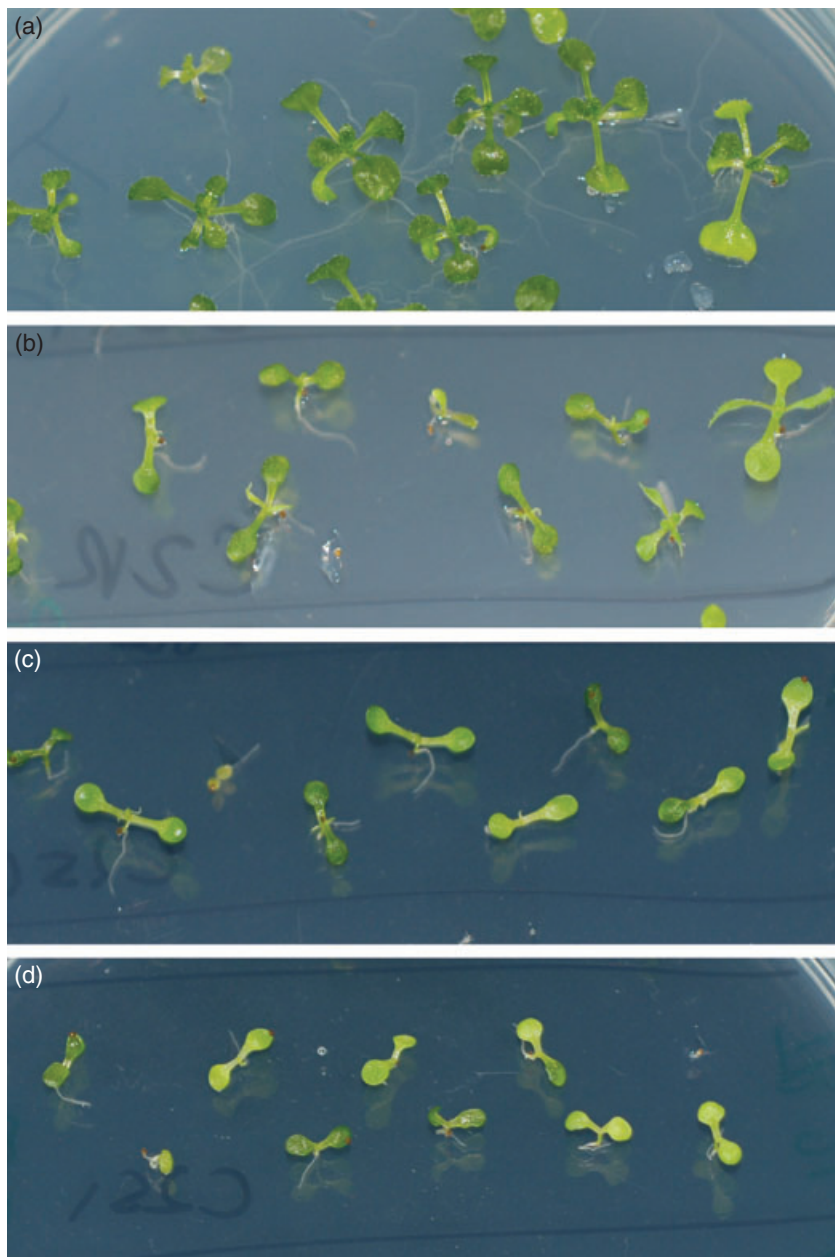
**Figure 2.** Zebularine treatment affects plant growth and development.
Arabidopsis seedlings grown for 14 days on zebularine-containing medium with (a) 0 μM, (b) 20 μM, (c) 40 μM or (d) 80 μM zebularine. Images were taken 14 days after sowing.

*Zebularine causes a dose-dependent and transient reduction of global 5-methyldeoxycytidine levels in plants*

To investigate the effect of the drug treatment on the overall levels of 5-methyldeoxycytidine (5-mdC), mock- and zebularine-treated plants were compared with plants in which DNA methylation was reduced by genetic means. Mutations in the *DDM*1 gene drastically decrease the level of 5-mdC (Jeddeloh *et al.*, 1999; Vongs *et al.*, 1993). Plants were germinated and grown for 21 days on media containing 0, 20, 40 or 80 μM zebularine prior to preparation of genomic DNA. Global 5-mdC levels were analysed as a percentage of 5-mdC in relation to total deoxycytidine (dC) levels using

cation exchange HPLC (Rozhon *et al.*, 2008). Mock-treated wild-type seedlings (accession Zürich) had 6.2% 5-mdC, whereas the level was reduced to 4.4% in *ddm*1-5 seedlings, which is in agreement with previously published values (Leutwiler *et al.*, 1984; Rozhon *et al.*, 2008). Levels of 5-mdC in zebularine-treated seedlings were also significantly decreased in a dose-dependent manner, ranging from 5.6, 5.1 to 4.0% in plants treated with 20, 40 and 80 μM zebularine, respectively (Figure 3a). Therefore, zebularine can induce significant hypomethylation similar to genetically achieved levels.

We also analysed global 5-mdC levels in DNA from the leaf tissue of adult plants grown for 8 weeks without an

**Figure 3**. Global levels of 5-methyldeoxycytidine (5-mdC) are reduced by zebularine treatment.
(a) Genomic DNA extracted from mock-treated Arabidopsis seedlings or seedlings grown on 20–80 μM zebularine were analysed in triplicate for 5-mdC content by HPLC. The 5-mdC levels were compared with *ddm*1-5 mutant seedlings and adult plants after 8 weeks' recovery. Zh, wild-type accession Zürich.
(b) Zebularine reduces the global level of 5-mdC in *Arabidopsis thaliana* accession Zürich even more than the same concentration of 5-azacyti-dine (5-azaC; 40 μM, same protocol).
(c) Zebularine reduces 5-mdC levels in *Medicago sativa*.
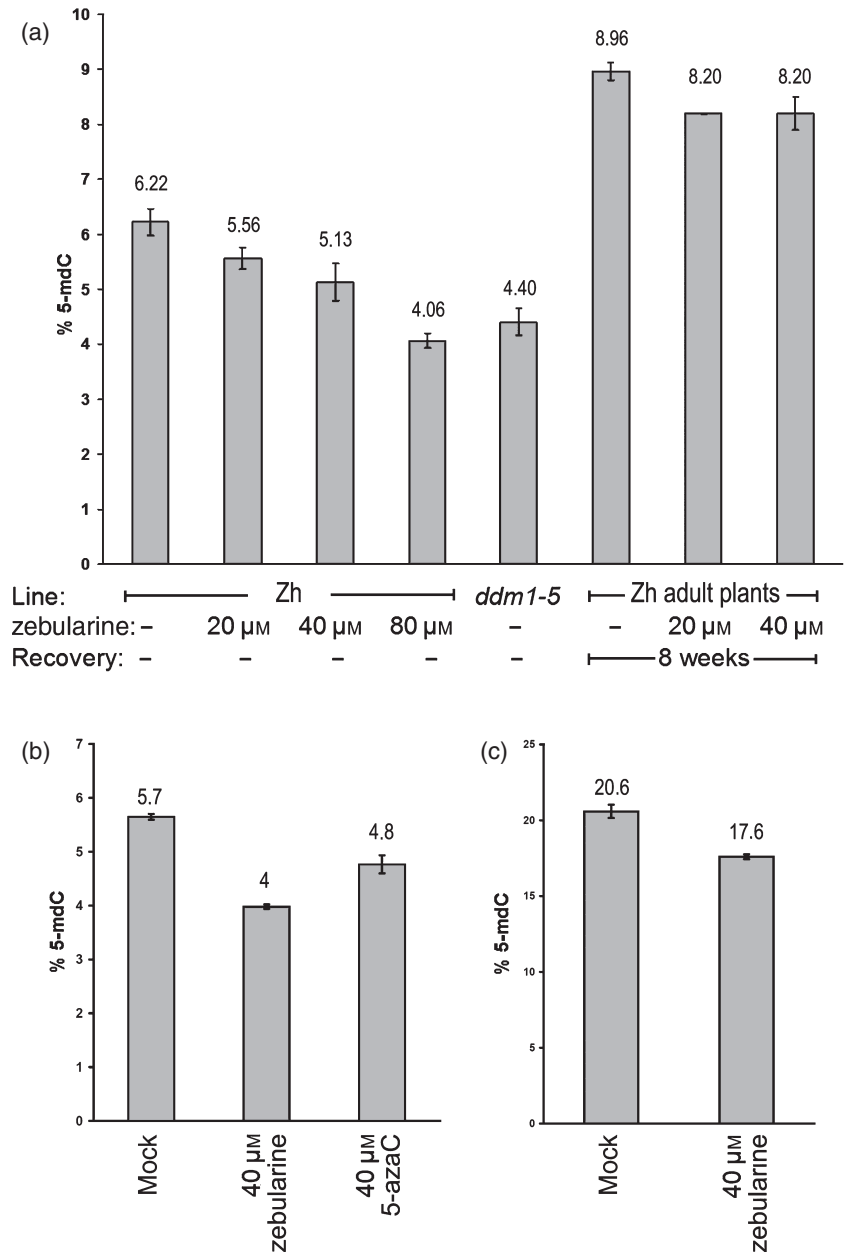


inhibitor, following the initial 21-day treatment with 0, 20 and 40 μM zebularine. DNA from all mature leaf samples had 1.4–1.6-fold more 5-mdC than seedlings, reflecting the developmental changes of DNA methylation levels previously described for untreated plants (Rozhon *et al.*, 2008; Ruiz-Garcia *et al.*, 2005). The difference between mock- and zebularine-treated adult plants decreased to insignificant values (Figure 3a), in agreement with the phenotypic recovery. Therefore, zebularine-induced reduction in 5-mdC levels, even at levels similar to genetically caused hypomethylation, is transient and can be overcome, at least globally, by growth in the absence of the drug.

To compare the efficiency of zebularine with the commonly applied but less stable DNA methylation inhibitor

5-aza, wild-type plants were germinated and grown for 21 days side-by-side on freshly prepared 0 or 40 μM zebularine- or 5-aza-containing media and analysed for the global 5-mdC levels as described. These were decreased in zebularine-treated plants to 4.0% (±0.04) and upon 5-aza treatment to 4.8% (±0.17) (Figure 3b). Therefore, zebularine is as efficient as, if not more so, than the commonly applied inhibitor 5-aza.

To test whether zebularine is effective in plant species other than *A. thaliana*, 5-mdC levels of *M. sativa* seedlings either mock-treated or treated with 40 μM zebularine for 1 week were analysed using the method described above. Mock-treated *Medicago* had 20.6% (±0.44) 5-mdC as previously reported (Rozhon *et al.*, 2008), whereas zebula-

rine-treated *Medicago* had only 17.6% (±0.16) 5-mdC (Figure 3c). This indicates that zebularine is also a potential inhibitor of DNA methylation in other plant species.

## Zebularine causes transient hypomethylation at transcriptionally inactive repeats

In order to elucidate whether the zebularine-induced DNA hypomethylation would affect different genomic regions in the same or in distinct ways, we conducted Southern blot experiments using methylation-sensitive restriction enzymes and sequence-specific probes homologous to different endogenous target sites known to be methylated. These included repetitive sequences such as *Athila*-related transcriptionally silent information (TSI) and 180-bp centromeric repeats. Both are highly methylated and either not expressed or practically not expressed in wild-type plants, but become hypomethylated and transcribed in *met*1 or *ddm*1 mutants (Mittelsten Scheid *et al.*, 1998; Steimer *et al.*, 2000; Vongs *et al.*, 1993). To distinguish DNA methylation at CG sites and CHG sites, we used the restriction enzyme *Hpa*II (sensitive to methylation at both cytosine residues in the recognition site CCGG) and its isoschizomere *Msp*I (limited only by ᵐCCGG; McClelland *et al.*, 1994).

As expected, repeat sequences from control plants were not cut by *Hpa*II and only weakly by *Msp*I, indicating strong methylation in both sequence contexts prior to drug treatment. Zebularine-treated plants showed DNA hypomethylation most prominently at CG sites of both TSI and 180-bp repeats, in a concentration-dependent manner (Figure 4a,b). The CHG sites were also affected, but to a lesser extent. Although the total content of 5-mdC in drug-treated plants was reduced to the same low level as in *ddm*1-5 plants, the hypomethylation of TSI and 180-bp repeats at CG and CHG sites was less pronounced than in the mutants. This indicates that the effects of zebularine are not biased towards demethylation of repetitive sequences, in contrast to the effect of the *ddm1* mutation (Vongs *et al.*, 1993).

While the restoration of DNA methylation patterns at repetitive regions can take several generations after outcrossing the *ddm*1 mutation (Kakutani *et al.*, 1996), methylation at TSI repeats is essentially restored in plants that were allowed to recover for 8 weeks after zebularine treatment (Figure 4a). The same was observed at 180-bp repeats, although prolonged exposure of the blots showed some minor remnants of demethylated repeats in recovered plants (Figure 4b).

## Zebularine causes dispersion of heterochromatic chromocentres but not complete depletion of 5-mdC

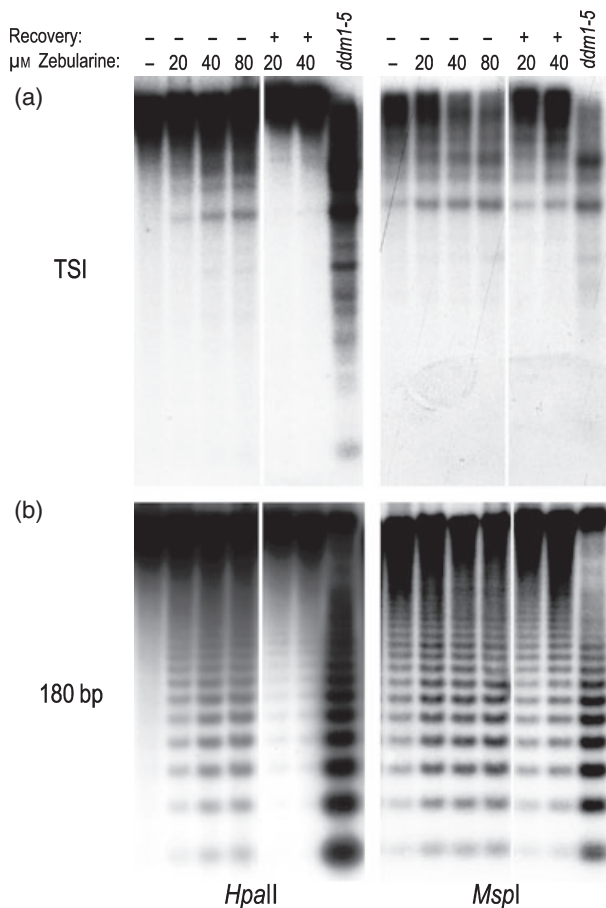Centromeric and pericentromeric repeats in Arabidopsis form heterochromatin that remains strongly condensed in



**Figure 4.** DNA methylation at repetitive sequences is decreased after zebularine treatment.
Genomic DNA from mock- or zebularine-treated plants and *ddm*1-5 mutants was digested with *Hpa*II and *Msp*I (sensitive to CG and CHG methylation, respectively) and hybridized to (a) transcriptionally silent information (TSI) and (b) 180-bp centromeric (pAL) repeats. Adult plants, recovered for 8 weeks after zebularine treatment, were also included.

interphase nuclei. These chromocentres (CCs) become decondensed and diffuse upon hypomethylation at centromeric repeats in *ddm*1 mutants (Probst *et al.*, 2003; Soppe *et al.*, 2002). Fluorescence *in situ* hybridization on nuclei from plants treated with 40 μM zebularine indeed contained less prominent and more dispersed CCs, as in *ddm*1 (Figure 5a–c), and these were significantly more frequent in zebularine-treated samples (25%) versus mock treatment (5%), and in a similar range as in *ddm*1 (34%) (Figure 5d). Thus, zebularine treatment causes similar changes in CC morphology as the *ddm*1 mutation.

While 5-mdC seems to be nearly erased from the residual condensed chromatin in *ddm*1, as seen upon immunostaining, the modification is still prominent at the remaining CCs in the drug-treated samples (Figure 5e–g). This is in accordance with the different degree of demethylation at the centromeric repeats seen at the molecular level for *ddm*1
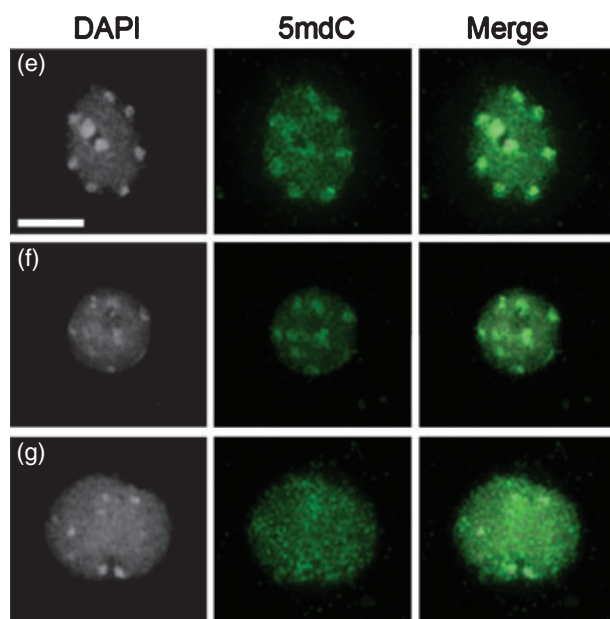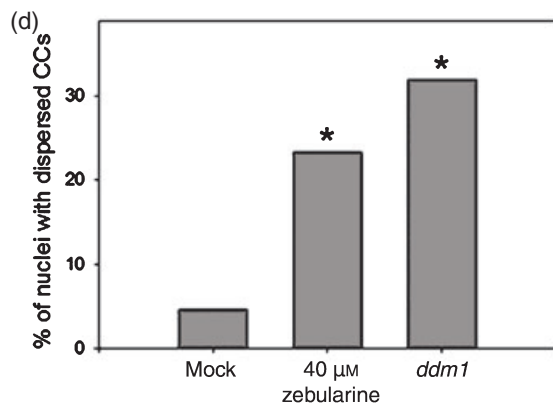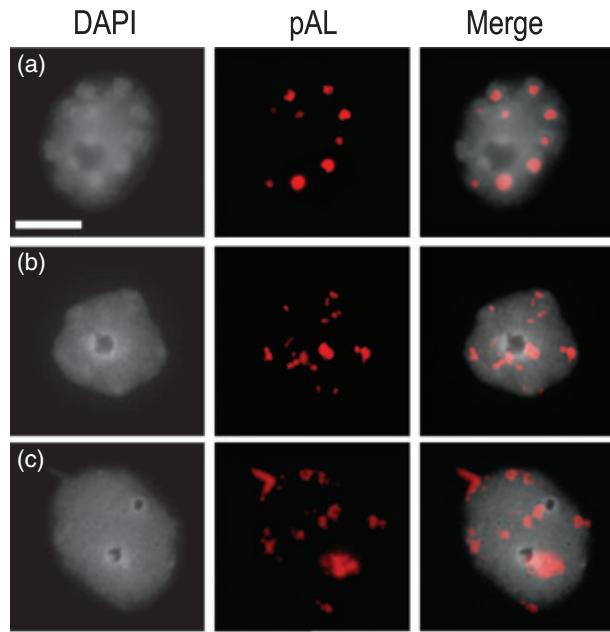
**Figure 5.** Morphology of centromeric repeats and distribution of 5-methyl-deoxycytosine (5-mdC) in zebularine-treated nuclei.

Cytological analysis by fluorescent *in situ* hybridization (FISH) with centromeric repeats (180 bp, pAL) revealed nuclei with either compact or dispersed signals, the first type representative of nuclei from mock-treated plants (a), the latter characteristic of nuclei from plants treated with 40 μM zebularine (b) or *ddm*1 mutants (c). (d) Nuclei with dispersed chromocentres are five-fold and seven-fold more abundant after treatment with 40 μM zebularine (*n* = 150) and in *ddm*1 mutant plant (*n* = 50) nuclei when compared with mock-treated nuclei (*n* = 150) (*t*-test, *$P$ < 0.001).

Immunolocalization of 5-mdC shows an unchanged distribution and signal intensity in (e) mock-treated and (f) 40 μM zebularine-treated nuclei, regardless of their dispersed chromocentres. (g) *ddm*1 nuclei display a strong reduction of DNA methylation at the chromocentres; however, gene body methylation is visible as uniform staining of euchromatin and seems not to be affected. Bars, 5 μm. DAPI, 4′,6-diamidino-2-phenylindole.

and zebularine treatment (Figure 4). However, the limited loss of methylation by zebularine apparently seems sufficient to loosen condensation of the CCs, and the presence of 5-mC immunofluorescence signals in CCs adds to the evidence that zebularine induces a rather unbiased loss of DNA methylation throughout the genome.

*Zebularine causes reactivation of transcriptionally inactive endogenous loci*

Perturbation of DNA methylation by genetic means or by inhibitors is frequently associated with transcriptional reactivation of otherwise hypermethylated sequences, such as repetitive endogenous sequences or some transgenes. Plant transposons are tightly regulated by the DNA methylation machinery to prevent replication and further spreading throughout the plant genome (Zilberman and Henikoff, 2004). Their transcription can serve as indicators for interference with methylation (Jeddeloh *et al.*, 1998, 1999; Kankel *et al.*, 2003). Therefore, we analysed plants grown on increasing dosages of zebularine for transcriptional activity of TSI and different transposons. Increasing amounts of zebularine led to a dose-dependent release of silencing at TSI loci and up-regulation of CACTA-like and MULE transposons as well as the LINE1-4 non-long terminal repeat (LTR) retrotransposon (Figure 6a,b). The expression of *ACTIN* and *TUBULIN8* was not affected by zebularine treatment (Figure 6b), allowing these genes to serve as loading controls.

Endogenous single-copy genes have also been reported to be regulated by DNA methylation, such as the imprinted *FWA* gene that is methylated in the promoter region and not expressed in vegetative plant tissues (Soppe *et al.*, 2000). However, *FWA* expression is induced in *ddm*1 and *met*1 mutants (Kakutani, 1997; Soppe *et al.*, 2000). We analysed *FWA* expression in zebularine-treated plants by quantitative RT-PCR and observed a dose-dependent increase in *FWA* mRNA levels after zebularine treatment. The highest dose resulted in a six-fold up-regulation compared with mock-treated plants (Figure 6c). Thus, zebularine treatment can
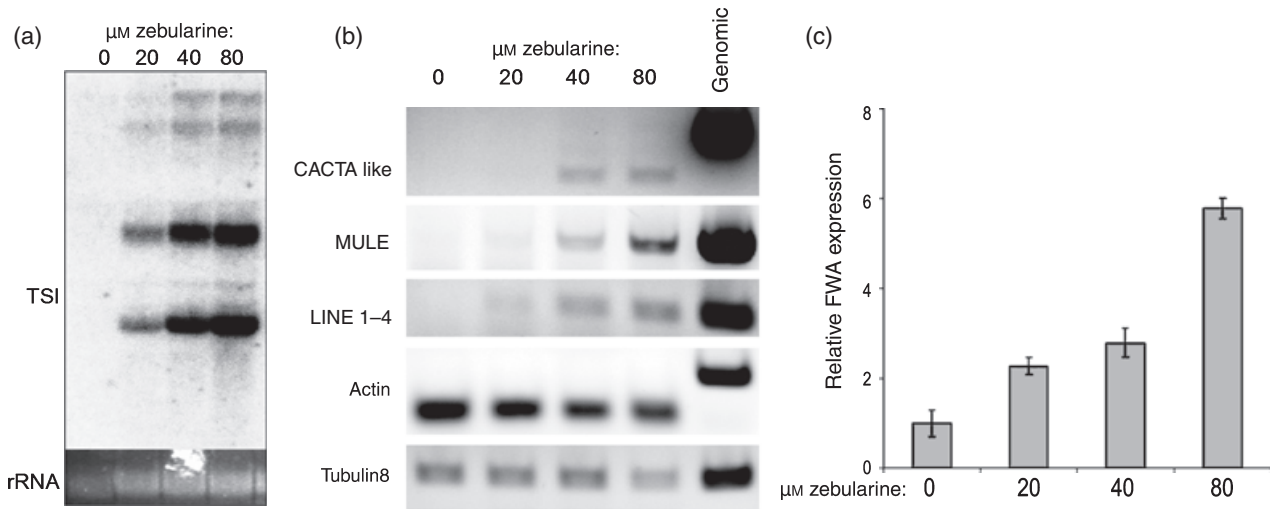
**Figure 6.** Zebularine-dependent reactivation of transcriptionally silenced genes.
(a) Northern blot analysis for transcriptionally silent information (TSI) mRNA accumulation after zebularine treatment.
(b) The RT-PCR assay for CACTA-like, MULE and LINE1-4 transposon reactivation after zebularine treatment. Actin and tubulin transcripts were used as loading controls.
(c) Abundance of *FWA* transcript in relation to Elongation Initiation Factor 4A (*EIF*4A) mRNA in pooled seedlings measured in triplicate by RT-qPCR.

induce transcriptional activity at repetitive and single-copy sequences that are otherwise hypermethylated and not expressed.

### Zebularine treatment affects DNA methylation of CG, CHG and CHH sites

The data described above indicated that the demethylating and transcriptionally reactivating effect of zebularine did not discriminate between the location of 5-mdC within repetitive sequences or single-copy genes. To further investigate whether the effect was also independent of the directly adjacent sequence context and whether zebularine inhibits all methyltransferases equally, we investigated the loss of DNA methylation after drug treatment by bisulphite conversion and sequencing. To focus the analysis on a sequence with a well-defined methylation pattern, we chose one of the short interspersed nucleotide element (SINE)-related direct repeats at the *FWA* gene, which is silent during the vegetative phase of Arabidopsis (Kinoshita *et al.*, 2007; Soppe *et al.*, 2000). Bisulphite sequencing can detect DNA methylation at every cytosine residue in a given sequence with high resolution. Bisulphite conversion was performed on DNA obtained from seedlings that were grown for 3 weeks on 80 μM zebularine, with mock-treated plants of the same age as controls. Total DNA methylation was reduced in zebularine-treated plants to 58.8% of all available sites, compared to 81.4% in untreated wild-type plants. The CHG and CHH methylation data published previously for the same sequence (http://epigenomics.mcdb.ucla.edu/DNA-meth/) (Cokus *et al.*, 2008) are slightly lower, probably reflecting an ecotype-dependent methylation polymor-

phism. However, zebularine treatment affected all sites: for CG from 98.3–90.3%, for CHG from 95–58.3% and for CHH from 75.3–50% (Figure 7a). With methylation in mock-treated plants set at 100%, the drug application reduced relative values by 8.1% (CG), 38.7% (CHG) and 33.6% (CHH). Thus, demethylation by zebularine appears to be unbiased with regard to the sequence context and seems to affect all methyltransferases.

Since zebularine was more effective than the *ddm*1 mutation with regard to global methylation, but induced less demethylation at repetitive sequences than the mutation, we asked whether the substantial methylation at coding regions of many genes would be affected. We extended the bisulphite sequence analysis to two genes that contain CG-specific gene body DNA methylation (Zilberman *et al.*, 2007) which is reduced in a *met*1 mutant background (Zhang *et al.*, 2006). A MutS DNA mismatch repair gene (At1g65070) and a RNA helicase (At3g06480) have 85.5% and 92.6% CG site-specific methylation, respectively, in mock-treated plants. After 80 μM zebularine treatment, these values are reduced by 23.6% and 19.4% CG methylation for At1g65070 and At3g06480, respectively (Figure 7b). Zebularine therefore induces hypomethylation at all types of sequences, in an unbiased manner and apparently in proportion to the degree of pre-existing methylation.

### Zebularine induces reactivation of transcriptionally inactive transgenic loci

Changes in epigenetic regulation are frequently analysed based on reporter genes whose expression can be visualized

**Figure 7.** Zebularine treatment reduces *FWA* promoter methylation and genic methylation. (a) Total and sequence context-specific DNA methylation determined by bisulphite sequencing of eight clones representing the *FWA* promoter and (b) CG methylation at two coding regions of genes with gene body methylation after treatment with 80 μM zebularine (8 and 12 clones, respectively).



(a)
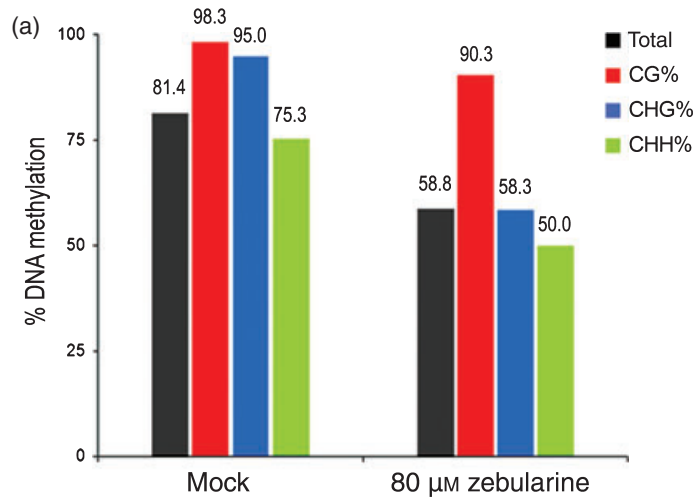
| (b) Gene | Region analyzed | #CG sites available | #CG sites methylated | Zebularine treatment | % CG methylation |
|---|---|---|---|---|---|
| *At1g65070* | 500 bp | 11 | 9 | Mock | 85.5% |
| | | | | 80 μM | 76.4% |
| *At3g06480* | 300 bp | 14 | 9 | Mock | 92.6% |
| | | | | 80 μM | 80.6% |

by enzymatic staining reactions or fluorescence. In plants, the β-glucuronidase reporter (GUS) and green fluorescent protein (GFP) are widely used reporter, and transgenic lines with transcriptionally silenced marker genes are available for both. TS-GUS (6b5/L2, (Morel *et al.*, 2000; Probst *et al.*, 2004)) and TS-GFP (L5, T. Blevins and F. Meins, Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland, pers. comm.) contain repetitive GUS or GFP genes, respectively, which had been shown previously to become reactivated in the background of mutants affecting DNA methylation and chromatin remodelling, such as *ddm*1-5, *met*1-3 or *mom*1-1 (Amedeo *et al.*, 2000; Morel *et al.*, 2000; T. Blevins, pers. comm.). To visualize reactivation by zebularine-induced DNA demethylation *in planta*, seedlings of lines TS-GUS and TS-GFP were grown for 21 days on plates containing zebularine prior to analysis for GUS and GFP expression. Mock-treated seedlings showed neither significant GUS staining nor GFP expression (Figure 8a,e), whereas the zebularine treatment released silencing of TS-GUS at concentrations of 20, 40 and 80 μM (Figure 8b–d). The TS-GFP plants, pre-treated with 40 μM zebularine, were also positive for transgene expression (Figure 8f).

The methylation inhibitor 5-aza had been shown to act synergistically in combination with trichostatin A (TSA), a histone de-acetylase inhibitor affecting gene silencing in animals and plants (Chen and Pikaard, 1997; Gartler and Goldman, 1994), although the interaction in plants is complex and can be antagonistic for certain target genes (Chang and

Pikaard, 2005). We therefore tested zebularine in combination with TSA. The TS-GUS and TS-GFP seeds were germinated on media with either 1.6 μM (0.5 μg ml⁻¹) TSA or 40 μM zebularine or both drugs at the same concentration as for the single treatments. Trichostatin A alone did not reactivate the silent reporter *GUS* gene even after 3 weeks of application (Figure 8h). A synergistic effect of TSA and zebularine was observed on plant growth and development, which were inhibited since seedlings treated with both drugs were much smaller than mock-, TSA- or zebularine-treated seedlings. However, the effect of the drug combination upon reporter gene expression seemed to be rather the opposite, because staining in TS-GUS plantlets was less intense than with zebularine treatment alone (Figure 8i). This might be due to the general growth inhibition that could reduce the potential for GUS and GFP expression, or indicate antagonistic effects between histone deacetylase inhibitors and 5-mdC inhibitors similar to those reported earlier (Chang and Pikaard, 2005).

## Discussion

Methylation of cytosine residues is the most frequent chemical modification of genomic plant DNA and is found in such amounts that the terminology of the 'fifth nucleotide' (Doerfler, 2006) is as justified in this kingdom as for mammalian DNA. 5-Methyldeoxycytidine is an important element of epigenetic regulation in plants, diverse with regard to sequence context, location at gene bodies or non-coding
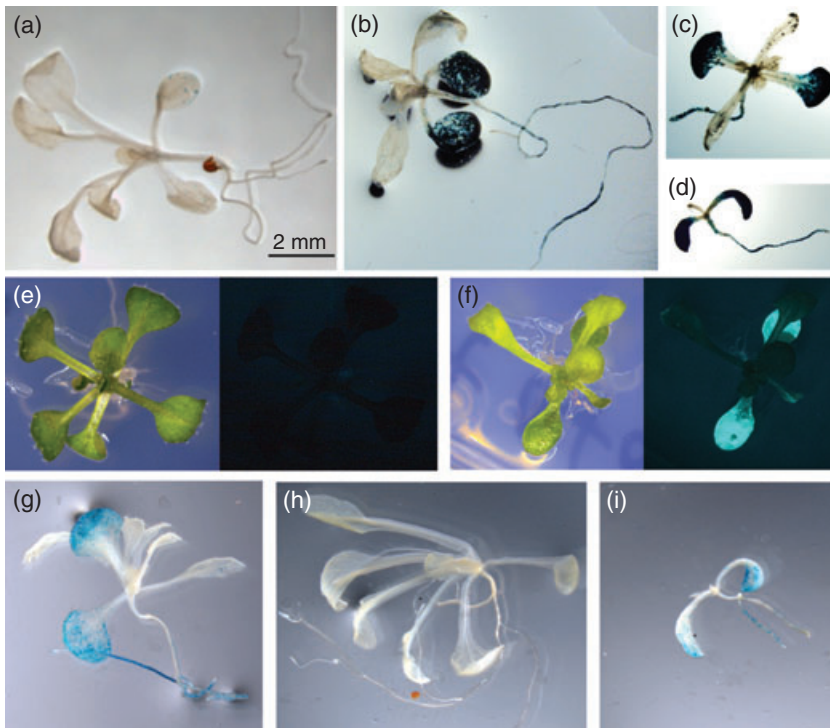
**Figure 8.** Effects of zebularine and/or trichostatin A on repetitive transcriptionally silent transgenes.

Transgenic lines with repetitive, transcriptionally silent GUS (TS-GUS) (a–d) genes are reactivated by (b) 20 μM, (c) 40 μM and (d) 80 μM zebularine. Additionally, a line containing transcriptionally silent GFP (TS-GFP; e, f) genes is reactivated by (f) 40 μM zebularine. Mock-treated plants (a) and (e) were used as controls. TS-GUS seedlings, 3 weeks old after treatment with (g) 40 μM zebularine, (h) 40 μM trichostatin A and (i) a combination of both.

regions, single-copy sequences or repeats. It is a stable, covalent modification, yet amenable to addition or removal by enzymatic activities or to passive loss upon replication or loss of functional methyltransferases. Specific methylation inhibitors are considered to be important tools for studying the biological role of DNA methylation, as apparent from the frequent use of the methylation inhibitor 5-aza (Lyko and Brown, 2005). However, this compound has an extremely short half-life in aqueous solutions such as plant growth media, not matching the long culture periods necessary for plant development. Therefore, to achieve reliable and reproducible general demethylation, the potential of zebularine, an agent with a similar mode of action to 5-aza but significantly better chemical stability (Cheng *et al.*, 2003; Zhou *et al.*, 2002), was explored in plant culture, and indeed, the presence of zebularine in the growth medium induced a significant, global reduction of 5-mdC in two plant species. In a direct comparison, zebularine caused an even slightly higher global demethylation than 5-aza, which can be due to better uptake, better integration or most likely due to higher stability in the plant culture medium. We showed that, in Arabidopsis, zebularine induced a non-discriminative and dosage-dependent reduction of 5-mdC. This offers certain advantages over the use of genetic mutants affecting DNA methylation only in certain sequence contexts such as $^{m}CG$, $^{m}CHG$ or $^{m}CHH$, or restricted to certain chromosomal regions and targets.

The preferential loss of methylation at centromeric regions in *ddm*1 mutant nuclei causes a significant decon-

densation and dispersion of the centromeric heterochromatin. The hypomethylation by zebularine is much less pronounced at centromeric repeats, as is apparent from molecular and cytological analysis. Nevertheless, *ddm*1 mutant nuclei and inhibitor-treated material showed a similar change in nuclear organization. This indicates that small changes in the methylation level are sufficient to interfere with the maintenance of the condensed state. Alternatively, the methylation status of other regions may contribute to condensation of heterochromatic regions, by recruiting interacting proteins or shaping larger complexes of nuclear organization. A direct or indirect effect of demethylation on nuclear organization at the chromosome level has also been observed for centromeres in polyploid wheat: the somatic association of homologous as well as homeologous centromeres was significantly reduced in xylem vessel cells upon treatment of roots with 5-aza (Vorontsova *et al.*, 2004).

Loss of DNA methylation upon genetic interference can become more drastic over several generations of inbreeding homozygous mutants (Kakutani *et al.*, 1996) or persist into subsequent generations even upon restitution of the methylation machinery after outcrossing with wild-type plants (Kakutani *et al.*, 1999). Data for application of 5-aza are not unambiguous. While there is a claim for heritable demethylation and morphological consequences in progeny of treated rice seedlings (Sano *et al.*, 1990), other studies have shown a transient effect (Kumpatla and Hall, 1998). Conversely, the demethylating effect of zebularine is transient,

since DNA methylation level and patterns are restored in somatic tissue formed after removal of the drug. This suggests that the blueprint for the methylation patterns is not fully removed. It could either reside in the residual methylation itself or in some other chromatin-associated information that may be erased by the mutations but not by zebularine. Extension of the methylation analysis to both strands of the same genomic template by hairpin bisulphite sequencing (Laird *et al.*, 2004) could permit investigation into how far the methylation is erased from both Cs at symmetric methylation sites. Together with pulsed application of zebularine-induced demethylation, this will allow an analysis of the pre-requisites and kinetics of remethylation.

The response of transcriptionally silenced targets to zebularine treatment was crucial to claim an equal or superior action of this drug. This has been proven for several endogenous indicators (centromeric repeats and transposons) and repetitive transgenic marker genes (TS-GUS, TS-GFP) as well as for protein-coding genes that are under transcriptional control of neighbouring low-copy repeats (*FWA*). Their dose-dependent reactivation after zebularine treatment seems to be directly connected with the dose-dependent demethylation. Interestingly, the three transposons included in our study respond in a similar way (although to different levels; Figure 6). This is not the case upon genetic interference with methylation: while retrotransposon LINE1-4 is significantly activated in a *ddm*1, *met*1 and *cmt*3 mutant background, the Mule transposon is not up-regulated in *cmt*3 (Lippman *et al.*, 2003). This is further evidence that zebularine discriminates less between different methylation types and targets. Data about release of these transposons from silencing by treatment with 5-aza are not available, since they were underrepresented on the microarrays used in the otherwise most comprehensive study of Chang and Pikaard (2005). However, a direct comparison of the two drugs in human cell culture indicated that both could reactivate a methylated gene relevant for cell adhesion and invasiveness, while 5-aza treatment (even at a much lower dose) additionally activated a latent virus (Rao *et al.*, 2007). This may indicate a different spectrum of action and allows a fine-tuned application of zebularine for specific experimental purposes.

## Experimental procedures

### Plant growth and chemical treatments

Cold-treated seeds were sterilized in 5% sodium hypochlorite and 0.05% Tween-80 for 6 min, washed and air-dried overnight. Sterilized seeds were sown and grown directly onto Petri dishes with agar-solidified germination medium containing zebularine (Sigma, http://www.sigmaaldrich.com/), 5-aza (Sigma) and/or TSA (Sigma) and grown for 21 days in growth chambers under 16-h light/8-h dark cycles at 21°C. Zebularine and 5-aza in aqueous solution or TSA dissolved in DMSO were added to the germination medium before solidifying at final concentrations of 20, 40 and 80 μM of zebularine, 40 μM 5-aza and 1.6 μM (0.5 μg ml$^{-1}$) of TSA. Plants were transferred to drug-free growth medium after 14 or 21 days for recovery.

### Nucleic acid isolation and gel-blot analysis

Seedlings were harvested as pools of 100 plantlets, shock-frozen in liquid nitrogen and homogenized by vortexing for 1 min using two or three ceramic spheres of diameter 1 cm. Rosette and stem leafs from three to five adult plants were harvested, shock-frozen in liquid nitrogen and homogenized. Homogenized plant tissue was subsequently used for DNA or RNA extraction using Phytopure (Amersham, http://www5.amershambiosciences.com/) or RNAeasy (Qiagen, http://www.qiagen.com/) kits, respectively.

For Southern blot analysis, 10 μg of genomic DNA was digested overnight with 1–2 U *Hpa*II or *Msp*I (MBI Fermentas, http://www.fermentas.com/). Subsequently, samples were electrophoretically separated on 1.2% TRIS–acetate–ethylenediamine tetraacetic acid [TAE; TRIS = 2-amino-2-(hydroxymethyl)-1,3-propanediol] agarose gels, depurinated for 10 min in 250 mM HCl, denatured for 30 min in denaturation solution containing 0.5 M NaOH and 1.5 M NaCl and neutralized twice in 0.5 M TRIS, 1.5 M NaCl and 1 mM EDTA at pH 7.2 for 15 min. For northern blot analysis, 10 μg of total RNA was denatured with 15% glyoxal and DMSO for 1 h at 50°C and separated using 1.4% agarose gels in 10 mM sodium phosphate buffer pH 7 in a Sea2000 circular flow electrophoresis chamber (Elchrom Scientific, http://www.elchrom.com/). DNA and RNA gels were blotted onto Hybond N+ (Amersham) membranes overnight with 20× SSC, washed and UV-crosslinked using a Stratalinker (Stratagene, http://www.stratagene.com/). Hybridization was performed as described by Church and Gilbert (1984). Radioactive (50 μCi) dCT-α-$^{32}$P (Amersham) labelled sequence-specific probes (TSI-A15 and pAL-180 bp) were synthesized from 25 ng of DNA using the Rediprime labelling kit (Amersham) and purified on G50 Probequant (Amersham) columns. Signals were detected with Phosphorimager Screens (Bio-Rad, http://www.bio-rad.com/) and scanned with a Molecular Imager FX (Bio-Rad).

### Cation-exchange high-pressure liquid chromatography

Total cytosine methylation was determined as described (Rozhon *et al.*, 2008). In short, 5 μg of genomic DNA was digested overnight at 37°C with 0.0025 U nuclease P1 and 0.5 U DNaseI in 20 mM acetic acid, 20 mM glycine, 5 mM MgCl$_2$, 0.5 mM ZnCl$_2$ and 0.2 mM CaCl$_2$, pH 5.3 in a total volume of 50 μl. Subsequently, 5 μl of 0.1 M NaOH and 1 U calf intestine alkaline phosphatase were added and the mixture incubated for a further 24 h. Samples were acidified by addition of 44 μl of 12 mM HCl prior to injection into the HPLC system equipped with a 125 × 4 mm Nucleosil 100-10 SA column (Macherey-Nagel, http://www.macherey-nagel.com/) preceded by a Valco 2 μm inline filter. The mobile phase consisted of 60 mM acetic acid and 15% acetonitrile, pH 4.8, with a constant flow rate of 1.5 ml min$^{-1}$. Ultraviolet detection was performed at 277 nm with a bandwidth of 10 nm with a PDA-100 photodiode array detector, and chromatograms were analyzed with Chromeleon 7 (Dionex, http://www.dionex.com/). All samples were analysed in technical triplicates and 5-mdC values were expressed as a percentage of total cytosine.

### Reverse transcription PCR and real-time PCR

Prior to reverse transcription, 30 μl RNA solution was treated with 5 U DNase I (MBI Fermentas), 0.4 U ribonuclease inhibitor (Rnasin)

and 4 μl of 10× DNase I buffer for 40 min at 37°C to remove residual DNA contamination in the RNA samples, extracted with phenol:chloroform (24:1) and subsequently ethanol-precipitated. Reverse transcription was performed on 1 μg of RNA with 0.2 μg random hexamer primers (MBI Fermentas) using 1 U RevertAid M-MuLV-RTase, RNaseH- (MBI Fermentas) at 42°C for 1.5 h. The cDNA thereby obtained was used for PCR and real-time PCR. Standard PCR was performed with True-Start Taq polymerase (Promega, http://www.promega.com/) and the following primers: CACTA-F: 5′-GGCTAGCTGTCCGACTCAATGACCT-3′, CACTA-R: 5′-CAGACATC-CTTTCCTTCAGCTTAGC-3′, MULE2-F: 5′-CTGTCCGCGAGTGTCA-TCAAGTAGC-3′, MULE2-R: 5′-GATACTTGTTGACAAGTGTTTAGC-AAGCC-3′, FWA-RTF: 5′-GTGTTAATGATCAAGATGGTGGAA-3′, FWA-RTR: 5′-AAGCTCGTACCTCTGTTCTTCAGT-3′, ActinF: 5′-TC-CCTCAGCACATTCCAGCAGAT-3′, ActinR: 5′-AACGATTCCTGGAC-CTGCCTCATC-3′, SN1F: 5′-ACTTAATTAGCACTCAAATTAAACAA AATAGT-3′, SN1R: 5′-TTTAAACATAAGAAGAAGTTCCTTTTTCATC-TAC-3′, EIF4A-F: 5′-ATCCAAGTTGGTGTGTTCTCC-3′ and EIF4A-R: 5′-GAGTGTCTCGAGCTTCCACTC-3′. Real-time PCR analysis was performed with the DyNAmo SYBRgreen kit (New England Biolabs, http://www.neb.com/) using a Rotorgene 3000 (Corbett, http://www.corbettlifescience.com/) lightcycler with data acquisition at 72°C to avoid signals from primer dimers. Ct values were analysed using Excel (Microsoft, http://www.microsoft.com/).

### In situ *GUS and GFP detection*

The GUS activity was detected by staining in 0.1 M sodium phosphate buffer pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 100 μg ml$^{-1}$ chloramphenicol, 2 mM potassium ferrocyanide, 2 mM potassium ferricyanide and 0.5 mg ml$^{-1}$ X-Gluc after 30-min vacuum infiltration and overnight incubation at 37°C. Subsequent washes with 70% ethanol at 37°C were performed in order to remove residual chlorophyll. All samples were analysed using a Leica MZ16FA binocular microscope with a Leica DFC300FX CCD camera (http://www.leica.com/). Images were acquired with Leica Application Suite and processed with Adobe Photoshop (Adobe, http://www.adobe.com/). Plants transgenic for TS-GFP were analysed under UV illumination with a Leica GFP1 filter (excitation 425/60 nm, emission barrier 480 nm) directly on plates.

### *Fluorescence* in situ *hybridization (FISH) and immunolabelling detection*

For the preparation of nuclei, 21-day-old plantlets were rinsed in 10 mM TRIS buffer pH 7.5, fixed by vacuum infiltration in 4% formaldehyde/TRIS buffer, rinsed in TRIS buffer, chopped in 500 μl chromosome isolation (CI) buffer (15 mM TRIS, 2 mM Na2EDTA, 0.5 mM spermin, 80 mM KCl, 20 mM NaCl, 15 mM beta-mercapto-ethanol, 0.1% Triton X-100, pH 7.5) and filtered through a 50-μm nylon mesh. Fifty microlitres of nuclei suspension was transferred onto microscope slides using Cytospin (560 **g** for 10 min). After centrifugation, slides were shortly rinsed in 1× PBS, transferred into 50% glycerol and stored at −20°C until use.

Immunolocalization of methylated cytosine was performed as described (Jasencakova et al., 2000) with minor modifications. In brief, slides were treated with pepsin (50 μg ml$^{-1}$ in 0.01 M HCl; Roche, http://www.roche.com/) at 38°C (1–2 min), post-fixed in 4% formaldehyde/2× SSC, denatured in 70% formamide/2× SSC at 80°C (2 min) and cooled in ice-cold 1× PBS. After blocking (5% BSA, 0.2% Tween 20, 4× SSC) at 37°C (30 min), the slides were incubated with primary monoclonal mouse-anti-5-methylcytosine (1:500, Eurogentec, http://www.eurogentec.com/) and secondary goat-anti-mouse-

Alexa488 (1:250, Molecular Probes, http://www.invitrogen.com/site/us/en/home/brands/Molecular-Probes.html) antibodies.

A biotin-labelled Arabidopsis centromeric repeat (pAL, 180 bp) probe for FISH was prepared from genomic DNA by PCR using primers pALU 5′-AGTCTTTGGCTTTGTGTCTT-3′ and pALR 5′-TGGACTTTGGCTACACCATG-3′. Slide pre-treatment and detection steps were performed as described (Pecinka et al., 2004). The probe was detected with subsequent avidin-Texas Red (1:1000, Vector Laboratories, http://www.vectorlabs.com/), goat-anti-avidin-biotin (1:200, Vector Laboratories) and again avidin-Texas Red (1:1000). The slides were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) [1 μg ml$^{-1}$ in Vectashield (Vector Laboratories)] and analysed using a Zeiss Axioplan 2 epifluorescence microscope. Monochromatic images were acquired with MetaVue (http://www.moleculardevices.com/pages/software/metavue.html) and processed with Adobe Photoshop (Adobe).

### Bisulphite conversion, sequencing and evaluation

After treatment with RNaseA and proteinase K, 1–2 μg of genomic DNA was digested overnight with *Bam*HI (MBI Fermentas). Subsequent bisulphite conversion was carried out using the Epitect Conversion Kit (Qiagen) and controlled for completion as described (Hetzl et al., 2007). Converted DNA was used for PCR amplification with the following primer pairs: FWA-L1: 5′-GGGTTTAGTGTT-TAYTTGTTTAAGG-3′, FWA-R4: 5′-TCTRATTRTCARTATCCCACAA-ATC-3′, At1g65070bsF: 5′-GTATYYGTGAGATGTGGTTATTAAAG-GTTG-3′, At1g65070bsR: 5′-CATCACATACAAATTAAATTAATAAT-ATCTATCCC-3′, At3g06480bsF: 5′-GAAGTAGTATAAATAYGAATA-AAGGTAAGTAATTTTG-3′ and At3g06480bsR: 5′-CTRAAACA-AACCCATCCTTATAACRCARTATATT-3′ (Zilberman et al., 2007). The PCR-amplified DNA was cloned using CloneJet or InsTAclone kits (MBI Fermentas) and transformed into DH5α cells (Invitrogen), sequenced by terminal labelling using BigDye Terminator v3.1 (Applied Biosystems, http://www.appliedbiosystems.com/) and read at http://www.vbc-genomics.com. The sequence information obtained was analysed with CyMATE (http://www.gmi.oeaw.ac.at/cymate; Hetzl et al., 2007) and Excel (Microsoft).

### References

Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and Paszkowski, J. (2000) Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. *Nature*, **405**, 203–206.

Beisler, J.A. (1978) Isolation, characterization, and properties of a labile hydrolysis product of the antitumor nucleoside, 5-azacytidine. *J. Med. Chem.* **21**, 204–208.

Chang, S. and Pikaard, C.S. (2005) Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA

methylation and histone deacetylation. *J. Biol. Chem.* **280**, 796–804.

Chen, Z.J. and Pikaard, C.S. (1997) Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in Brassica. *Proc. Natl Acad. Sci. USA*, **94**, 3442–3447.

Cheng, J.C., Matsen, C.B., Gonzales, F.A., Ye, W., Greer, S., Marquez, V.E., Jones, P.A. and Selker, E.U. (2003) Inhibition of DNA methylation and reactivation of silenced genes by zebularine. *J. Natl Cancer Inst.* **95**, 399–409.

Christman, J.K. (2002) 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*, **21**, 5483–5495.

Church, G.M. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M. and Jacobsen, S.E. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **452**, 215–219.

Constantinides, P.G., Jones, P.A. and Gevers, W. (1977) Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. *Nature*, **267**, 364–366.

Dieguez, M.J., Bellotto, M., Afsar, K., Mittelsten Scheid, O. and Paszkowski, J. (1997) Methylation of cytosines in nonconventional methylation acceptor sites can contribute to reduced gene expression. *Mol. Gen. Genet.* **253**, 581–588.

Doerfler, W. (2006) The almost-forgotten fifth nucleotide in DNA: an introduction. *Curr. Top. Microbiol. Immunol.* **301**, 3–18.

Finnegan, E.J., Peacock, W.J. and Dennis, E.S. (1996) Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. *Proc. Natl Acad. Sci. USA*, **93**, 8449–8454.

Gartler, S.M. and Goldman, M.A. (1994) Reactivation of inactive X-linked genes. *Dev. Genet.* **15**, 504–514.

Ghoshal, K. and Bai, S. (2007) DNA methyltransferases as targets for cancer therapy. *Drugs Today (Barc)*, **43**, 395–422.

Herranz, M., Martin-Caballero, J., Fraga, M.F., Ruiz-Cabello, J., Flores, J.M., Desco, M., Marquez, V. and Esteller, M. (2006) The novel DNA methylation inhibitor zebularine is effective against the development of murine T-cell lymphoma. *Blood*, **107**, 1174–1177.

Hetzl, J., Foerster, A.M., Raidl, G. and Mittelsten Scheid, O. (2007) CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51**, 526–536.

Jasencakova, Z., Meister, A., Walter, J., Turner, B.M. and Schubert, I. (2000) Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. *Plant Cell*, **12**, 2087–2100.

Jeddeloh, J.A., Bender, J. and Richards, E.J. (1998) The DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis. *Genes Dev.* **12**, 1714–1725.

Jeddeloh, J.A., Stokes, T.L. and Richards, E.J. (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* **22**, 94–97.

Kakutani, T. (1997) Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana. *Plant J.* **12**, 1447–1451.

Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K. and Richards, E.J. (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl Acad. Sci. USA*, **93**, 12406–12411.

Kakutani, T., Munakata, K., Richards, E.J. and Hirochika, H. (1999) Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of Arabidopsis thaliana. *Genetics*, **151**, 831–838.

Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L. and Richards, E.J. (2003) Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics*, **163**, 1109–1122.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J., Koornneef, M. and Kakutani, T. (2007) Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J.* **49**, 38–45.

Klug, M. and Rehli, M. (2006) Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. *Epigenetics*, **1**, 127–130.

Kumpatla, S.P. and Hall, T.C. (1998) Longevity of 5-azacytidine-mediated gene expression and re-establishment of silencing in transgenic rice. *Plant Mol. Biol.* **38**, 1113–1122.

Laird, C.D., Pleasant, N.D., Clark, A.D., Sneeden, J.L., Hassan, K.M., Manley, N.C., Vary, J.C. Jr, Morgan, T., Hansen, R.S. and Stoger, R. (2004) Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc. Natl Acad. Sci. USA*, **101**, 204–209.

Leutwiler, L., Hough-Evans, B. and Meyerowitz, E. (1984) The DNA of Arabidopsis thaliana. *Mol. Gen. Genet.* **194**, 15–23.

Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.

Lippman, Z., May, B., Yordan, C., Singer, T. and Martienssen, R. (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1**, E67.

Lyko, F. and Brown, R. (2005) DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. *J. Natl Cancer Inst.* **97**, 1498–1506.

Marquez, V.E., Barchi, J.J. Jr, Kelley, J.A., Rao, K.V., Agbaria, R., Ben-Kasus, T., Cheng, J.C., Yoo, C.B. and Jones, P.A. (2005) Zebularine: a unique molecule for an epigenetically based strategy in cancer chemotherapy. The magic of its chemistry and biology. *Nucleosides Nucleotides Nucleic Acids*, **24**, 305–318.

Mathieu, O., Yukawa, Y., Sugiura, M., Picard, G. and Tourmente, S. (2002) 5S rRNA genes expression is not inhibited by DNA methylation in Arabidopsis. *Plant J.* **29**, 313–323.

McClelland, M., Nelson, M. and Raschke, E. (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* **22**, 3640–3659.

Mittelsten Scheid, O., Afsar, K. and Paszkowski, J. (1998) Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. *Proc. Natl Acad. Sci. USA*, **95**, 632–637.

Morel, J.B., Mourrain, P., Beclin, C. and Vaucheret, H. (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr. Biol.* **10**, 1591–1594.

Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J. and Schubert, I. (2004) Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. *Chromosoma*, **113**, 258–269.

Probst, A.V., Fransz, P.F., Paszkowski, J. and Mittelsten Scheid, O. (2003) Two means of transcriptional reactivation within heterochromatin. *Plant J.* **33**, 743–749.

Probst, A.V., Fagard, M., Proux, F. *et al.* (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.

**Rao, S.P., Rechsteiner, M.P., Berger, C., Sigrist, J.A., Nadal, D. and Bernasconi, M.** (2007) Zebularine reactivates silenced E-cadherin but unlike 5-Azacytidine does not induce switching from latent to lytic Epstein-Barr virus infection in Burkitt's lymphoma Akata cells. *Mol. Cancer*, **6**, 3.

**Ronemus, M.J., Galbiati, M., Ticknor, C., Chen, J. and Dellaporta, S.L.** (1996) Demethylation-induced developmental pleiotropy in Arabidopsis. *Science*, **273**, 654–657.

**Rozhon, W., Baubec, T., Mayerhofer, J., Mittelsten Scheid, O. and Jonak, C.** (2008) Rapid quantification of global DNA methylation by isocratic cation exchange high-performance liquid chromatography. *Anal. Biochem.* **375**, 354–360.

**Ruiz-Garcia, L., Cervera, M.T. and Martinez-Zapater, J.M.** (2005) DNA methylation increases throughout Arabidopsis development. *Planta*, **222**, 301–306.

**Sano, H., Kamada, I., Youssefian, S., Katsumi, M. and Wabiko, H.** (1990) A single treatment of rice seedlings with 5-azacytidine induces heritable dwarfism and undermethylation of genomic DNA. *Mol. Gen. Genet.* **220**, 441–447.

**Santi, D.V., Garrett, C.E. and Barr, P.J.** (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell*, **33**, 9–10.

**Scott, S.A., Lakshimikuttysamma, A., Sheridan, D.P., Sanche, S.E., Geyer, C.R. and DeCoteau, J.F.** (2007) Zebularine inhibits human acute myeloid leukemia cell growth in vitro in association with p15INK4B demethylation and reexpression. *Exp. Hematol.* **35**, 263–273.

**Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M. and Peeters, A.J.** (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell*, **6**, 791–802.

**Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I. and Fransz, P.F.** (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J.* **21**, 6549–6559.

**Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Mittelsten Scheid, O. and Paszkowski, J.** (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. *Plant Cell*, **12**, 1165–1178.

**Vongs, A., Kakutani, T., Martienssen, R.A. and Richards, E.J.** (1993) Arabidopsis thaliana DNA methylation mutants. *Science*, **260**, 1926–1928.

**Vorontsova, M., Shaw, P., Reader, S. and Moore, G.** (2004) Effect of 5-azacytidine and trichostatin A on somatic centromere association in wheat. *Genome*, **47**, 399–403.

**Weber, H., Ziechmann, C. and Graessmann, A.** (1990) In vitro DNA methylation inhibits gene expression in transgenic tobacco. *EMBO J.* **9**, 4409–4415.

**Yoo, C.B. and Jones, P.A.** (2006) Epigenetic therapy of cancer: past, present and future. *Nat Rev Drug Discov.* **5**, 37–50.

**Yoo, C.B., Cheng, J.C. and Jones, P.A.** (2004) Zebularine: a new drug for epigenetic therapy. *Biochem. Soc. Trans.* **32**, 910–912.

**Zhang, X., Yazaki, J., Sundaresan, A.** *et al.* (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*, **126**, 1189–1201.

**Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J. and Hornby, D.P.** (2002) Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J. Mol. Biol.* **321**, 591–599.

**Zilberman, D. and Henikoff, S.** (2004) Silencing of transposons in plant genomes: kick them when they're down. *Genome Biol.* **5**, 249.

**Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. and Henikoff, S.** (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**, 61–69.