DNA Research **21**, 141–151, (2014) doi:10.1093/dnares/dst046
Advance Access publication on 7 November 2013

C^mCGG Methylation-Independent Parent-of-Origin Effects on Genome-Wide Transcript Levels in Isogenic Reciprocal F1 Triploid Plants

Mark T.A. Donoghue[1,†,‡], Antoine Fort[1,†], Rachel Clifton[1], Xu Zhang[2,¶], Peter C. McKeown[1], M.L. Voigt-Zielinksi[1,‖], Justin O. Borevitz[2,§], and Charles Spillane[1,*]

Genetics and Biotechnology Laboratory, Plant and AgriBiosciences Research Centre (PABC), School of Natural Sciences, Aras de Brun, National University of Ireland, Galway (NUI Galway), Ireland[1] and Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA[2]

*To whom correspondence should be addressed: E-mail: charles.spillane@nuigalway.ie

Edited by Dr Satoshi Tabata
(Received 29 July 2013; accepted 9 October 2013)

Abstract

Triploid F1 hybrids generated via reciprocal interploidy crosses between genetically distinct parental plants can display parent-of-origin effects on gene expression or phenotypes. Reciprocal triploid F1 isogenic plants generated from interploidy crosses in the same genetic background allow investigation on parent-of-origin-specific (parental) genome-dosage effects without confounding effects of hybridity involving heterozygous mutations. Whole-genome transcriptome profiling was conducted on reciprocal F1 isogenic triploid (3x) seedlings of *A. thaliana*. The genetically identical reciprocal 3x genotypes had either an excess of maternally inherited 3x(m) or paternally inherited 3x(p) genomes. We identify a major parent-of-origin-dependent genome-dosage effect on transcript levels, whereby 602 genes exhibit differential expression between the reciprocal F1 triploids. In addition, using methylation-sensitive DNA tiling arrays, constitutive and polymorphic CG DNA methylation patterns at CCGG sites were analysed, which revealed that paternal-excess F1 triploid seedling C^mCGG sites are overall hypermethylated. However, no correlation exists between C^mCGG methylation polymorphisms and transcriptome dysregulation between the isogenic reciprocal F1 triploids. Overall, our study indicates that parental genome-dosage effects on the transcriptome levels occur in paternal-excess triploids, which are independent of C^mCGG methylation polymorphisms. Such findings have implications for understanding parental effects and genome-dosage effects on gene expression and phenotypes in polyploid plants.

Key words: triploid; DNA methylation; parent-of-origin effect; polyploidy; epigenetic

1. Introduction

Changes in gene dosage at the whole-genome, chromosomal or segmental levels can elicit phenotypic and gene expression effects associated with dosage-sensitive genes.[1] Polyploidization events increase the dosage of all loci, including structural and regulatory loci, controlling traits that may be genome dosage sensitive. Due to the importance of polyploidy to plant evolution and crop breeding, many aspects of gene regulation in polyploids require elucidation,[2,3] including how dosage effects and other epigenetic changes are triggered and maintained in both allo- and autopolyploids.[4,5]

A range of studies in allopolyploid plant genomes have revealed rapid epigenetic changes, including alteration in cytosine methylation patterns, rapid silencing of ribosomal RNA and protein-coding genes, and

† These authors contributed equally to this study.
‡ Present address: Cold Spring Harbor Laboratory (CSHL), 1 Bungtown Road, Cold Spring Harbor, NY, USA.
¶ Present address: Department of Medicine, Comprehensive Sickle Cell Center, University of Illinois at Chicago, Chicago, IL, USA.
§ Present address: School of Biology, College of Medicine, Biology and Environment, Australian National University, Canberra, Australia.
‖ Present address: Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, 1030 Vienna, Austria.

de-repression of dormant transposable elements.[6,7] However, allopolyploid genomes are genetic hybrids, where both ploidy (genome dosage) and hybridity (mutational differences) occur in concert, making it difficult to disentangle genetic effects from genome (and gene)-dosage effects.

In contrast, the generation of autopolyploids in the same genetic background provides a model system for the analysis of genome-dosage effects (*sensu strictu*) in the absence of mutational differences between lines of different ploidy. In autopolyploids, there have been a range of studies on genome-dosage effects on phenotypes in maize[8,9] and *A. thaliana*.[10–12] Changes in genome dosage in autopolyploids, and also in individual gene dosage, have been shown to modify epigenetic silencing in plants.[9,13] The majority of studies on gene expression changes in autopolyploids conducted, to date, have either focused on a limited number of genes[12,14,15] or found few gene (transcript) expression level changes at the whole-genome level between diploids and tetraploids of *A. thaliana*.[16,17]

In addition, little is known regarding the extent of DNA methylation changes associated with autopolyploidy in *A. thaliana* or the functional effects of DNA methylation polymorphisms on gene expression or phenotypic changes. In diploids, it has been previously shown in *A. thaliana* methyltransferase mutants (*drm1*, *drm2*, *cmt3* and *met1*) that loss of methylation can lead to upregulation of genes,[18–20] and conversely that increased methylation can lead to downregulation in *A. thaliana* hybrids.[20]

In interploidy 2x × 4x crosses, genome-dosage effects can occur in a parent-of-origin-dependent or -independent manner, depending on whether the two different types of reciprocal F1 triploids (i.e. 2m:1p versus 1m:2p) display different phenotypes. While parent-of-origin-dependent genome-dosage effects have been observed on phenoypes in maize and *A. thaliana*, little is known regarding parent-of-origin-dependent genome-dosage effects on gene expression in reciprocal F1 triploids. The generation of reciprocal F1 triploids in the same genetic background provides a system for the identification and analysis of genome-dosage and other epigenetic parental effects on phenotypes and gene expression, which are not due to mutational differences between the reciprocal F1 triploids.

In this study, we have used isogenic reciprocal F1 triploids to demonstrate a major parent-of-origin-dependent (i.e. parental) genome-dosage-dependent effect on transcript levels in paternal genome excess F1 triploids. We also demonstrate that this novel parental genome-dosage effect in the F1 triploids is C^mCGG methylation-independent at the whole-genome level. This suggests that the paternally and maternally inherited chromosome sets in autopolyploid plants may be epigenetically different, due to parental genome-dosage effects that can affect transcript levels in a C^mCGG methylation-independent manner.

2. Materials and methods

2.1. Plant materials

Col-0 diploid (2x) and tetraploid (4x) seeds, selfed for at least two generations after colchicine treatment, were the kind gift of Luca Comai (University of Washington). Maternal-excess triploids (3x(m)) and paternal-excess triploids (3x(p)) were generated by manually crossing emasculated diploid or tetraploid flowers with diploid or tetraploid pollen under a Leica MZ6 dissection microscope using Dumostar No. 5 tweezers. Sterilized seeds were sown on 0.5X MS (Murashige and Skoog) media and grown in a Percival growth chamber (16 h light and 8 h darkness). Seedlings were harvested at the two true leaf stage (Boyes standard 1.02[21]) for subsequent analysis. The ploidy of the resulting crosses was verified by flow cytometry using a Partec Ploidy Analyzer, with CyStain UV Precise P (Partec) reagents, following manufacturer's instructions.

2.2. Sample preparation and microarray hybridization

RNA from four biological replicates per ploidy level, with 20 seedlings per replicate, was extracted using the QIAGEN RNeasy Plant Mini Kit (#74 903). mRNA was purified with QIAGEN Oligotex mRNA Mini Kits (#70 022) using 25 μg of initial RNA, and used for double-stranded cDNA synthesis (SuperScript Double-Stranded cDNA synthesis #11917-010). After RNAse treatment (Epicentre RNAseH) and dscDNA purification (QIAGEN Qiaquick PCR purification kit), samples were labelled using Invitrogen BioPrime DNA labelling, and processed. For methylation analysis, DNA was extracted using the QIAGEN DNeasy Plant Mini Kit (#69 104), and 300 ng of DNA digested by 20 units of *Mse*1 and 10 units of either *Hpa*II or *Msp*I (New England Biolabs) at 37°C for 16 h. After ethanol precipitation of digested DNA, the samples were labelled with Invitrogen BioPrime DNA labelling. All kits were used according to manufacturer's instructions. Transcriptome and methylome analyses were performed using a custom whole-genome, SNP-tiling array (AtSNPtile1).[22]

2.3. Tiling array and small RNA data analysis

The tiling array used was the AtSNPtile1-tiling array which contains 1.4 million unique probes tiled along both strands of the entire *A. thaliana* genome at 35 bp resolution. The tiling probes include all unique features with good hybridization quality on the *A. thaliana*-tiling array 1.0 (Affymetrix). The analysis of the tiling array data for the detection of indels,

gene expression and DNA methylation differences, including the validation of gene expression difference (by qRT-PCR) and correlation between DNA methylation, small RNA and gene expression is described in Supplementary Data.

3. Results and discussion

3.1. Generation of isogenic F1 triploid plants

Reciprocal crosses of tetraploid (4x) and diploid (2x) parental lines in the same accession background can generate viable reciprocal F1 triploids in *A. thaliana* that are genetically identical, i.e. they are isogenic (Fig. 1). Such reciprocal F1 triploids provide an ideal model system to investigate parent-of-origin-specific (parental) effects on gene expression and other phenotypes.[23] To identify any parent-of-origin-specific genome-dosage effects on gene expression between isogenic reciprocal F1 triploids of *A. thaliana* (accession Col-0), microarray profiling was performed on *A. thaliana* seedlings. The reciprocal F1 triploids were generated from neo-tetraploid (F2) plants (in the Col-0 accession background) that were reciprocally crossed in both parental directions (as either pollen or ovule donor) to the diploid progenitor line. Paternal-excess F1 triploids (3x(p)), each containing two copies of the paternal genome and one copy of the maternal genome, were generated by fertilizing emasculated diploid flowers with pollen from the tetraploid (i.e. a 2x × 4x cross). In contrast, emasculated tetraploid flowers were fertilized with pollen from the diploids (i.e. a 4x × 2x cross) to generate the reciprocal maternal-excess F1 triploids (3x(m)). The comparison of these genetically identical 3x(p) and 3x(m) plants formed the basis of our analysis of parent-of-origin genome-dosage effects on the transcriptome.

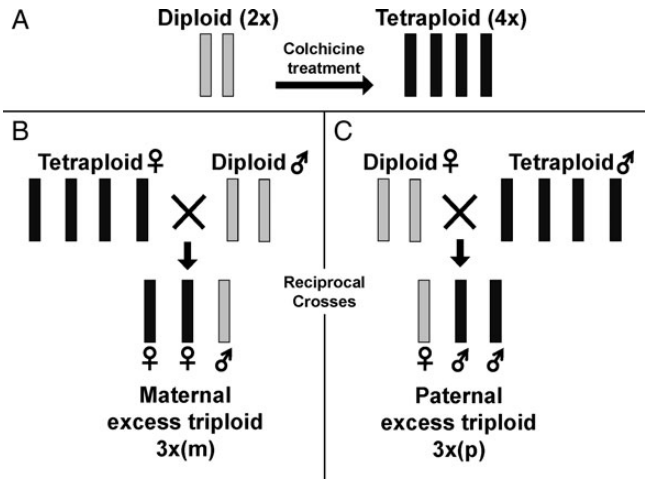

Figure 1. Generation of isogenic F1 reciprocal triploid plants. (A) Generation of tetraploid *A. thaliana* Col-0 using colchicine treatment. (B and C) Generation of maternal- and paternal-excess reciprocal F1 triploids by crossing diploid and tetraploid *A. thaliana* plants.

3.2. High-density tiling array analysis confirms that isogenic reciprocal F1 triploids generated from interploidy crosses in A. thaliana are genetically identical

In a selfing species, such as *A. thaliana*, the crossing of diploid and tetraploid plants that are in the same genetic background should result in the generation of F1 triploid plants that contain the same DNA sequence (i.e. are isogenic in terms of DNA sequence) as the parental lines. Such isogenic ploidy series systems allow us to test for strict genome-dosage effects on gene expression and other phenotypes. In addition, the generation of isogenic F1 triploids provides a model system for detecting parent-of-origin-specific genome-dosage effects. To confirm that reciprocal F1 triploids generated from interploidy crosses are indeed isogenic, we hybridized genomic DNA to the tiling arrays and screened for any evidence of single feature polymorphisms (SFPs)[24] by comparing individual probe intensities across the reciprocal F1 triploids. No SFPs were detected in the isogenic F1 triploids.

As this approach may not have detected longer indels spanning several tiling array probes, we also compared the two triploid datasets using a segmentation algorithm.[25] Only one potential indel (Chr2: 13 390 565–13 391 067) was predicted using this approach. When we subsequently tested this by PCR using primers designed to span the location of the putative indel, no such indel could be detected (Supplementary Fig. S1), indicating that this was a false positive from the segmentation algorithm. Overall, this analysis confirmed that the reciprocal F1 triploid plants are genetically isogenic and hence, provide a robust platform for the investigation of parent-of-origin-specific genome-dosage effects in plants.

3.3. Isogenic reciprocal F1 triploid plants display epigenetic parent-of-origin-specific genome-dosage effects on transcript levels

Having confirmed that the reciprocal F1 triploids generated from reciprocal interploidy crosses were truly isogenic, the tiling arrays were used to interrogate the expression levels of 25 703 genes in seedlings. Despite the fact that the reciprocal F1 triploids are genetically identical at the DNA sequence level, 602 genes were found to be differentially expressed between the reciprocal F1 triploids [using a false discovery rate (FDR) of $4.74 \times 10^{e-3}$]. All the 602 differentially

expressed genes were identified as having fold changes greater than 2.5 (Supplementary Table S1). All these 602 genes were upregulated in the paternal-excess 3x(p) F1 triploid compared with the maternal-excess 3x(m) F1 triploid. The upregulation of gene expression levels in 3x(p) versus 3x(m) F1 triploids was validated by qRT-PCR for 15 of 20 genes tested (75%) (Supplementary Fig. S2).

Overall, our results using isogenic F1 triploids provide the first evidence of widespread parent-of-origin-specific genome-dosage effects on gene expression in triploid vegetative plant tissues. While parent-of-origin-specific expression (e.g. due to genomic imprinting[26]) has been detected in triploid endosperm tissues,[27–29] previous studies have found no evidence for parent-of-origin effects on gene expression in diploid vegetative tissues.[22]

3.4. *The 602 genes subject to parent-of-origin-specific genome-dosage effects are enriched for stress-response genes*

To investigate the biological processes associated with the altered transcriptome of the 3x(p) triploid F1 seedlings, we screened for Gene Ontology (GO) terms enriched among the 602 upregulated genes by conditional hypergeometric tests.[30] In total, 872 biological processes (BPs), 546 molecular functions and 230 cellular components (CC) were tested for this analysis. The BP analysis discovered a significant parent-of-origin-specific genome-dosage effect on stress-response genes, with several stress-response terms significantly over-represented in the genes upregulated in the 3x(p) F1 triploid, including both biotic and abiotic stress responses (Supplementary Table S2).

To further investigate the enrichment for both abiotic and biotic stress-response genes in the 3x(p) triploid, the AtGenExpress abiotic stress, pathogen infection, growth conditions, hormone and chemical treatment datasets were also interrogated. For this analysis, a list of all differentially expressed genes was identified for each treatment–tissue–time-point combination in the AtGenExpress datasets, and these AtGenExpress lists were then compared with the 602 up-regulated genes in the 3x(p) triploid to identify significant overlaps. Using this approach, for each abiotic stress, tested (cold, salt, heat, osmotic, genotoxic, wounding, oxidative, drought and UV) significant overlaps were detected with the 602 genes identified in the 3x(p) triploid (Supplementary Table S3). In addition, significant overlaps were also detected between the 602 3x(p) genes and biotic stress-response gene sets. These included genes responsive to several pathogens, including *Pseudomonas*, potato blight, virulent, avirulent, Type III secretion system-deficient and non-host bacterial pathogens, bacterial derived elicitors [LPS, HrpZ, Flg22 and oomycete (NPP1)] and mildew infection (Supplementary Table S3). While genome-dosage effects on transcript levels have been observed between diploid and tetraploid *A. thaliana*,[16,17,31,32] and tetraploid lines observed to have increased salinity tolerance,[33] to our knowledge, this is the first time that a parent-of-origin-specific genome-dosage effect on transcript levels of biotic and abiotic stress-response genes has been demonstrated.

3.5. *Paternal-excess isogenic F1 triploids display elevated ^{m}CG methylation at CCGG sites*

It is commonly considered that genome-wide transcript expression is correlated with gene cytosine methylation,[18,34,35] which, in plants, most commonly occurs in the CG context.[36] To determine whether the genome-wide distribution of CG DNA methylation differed between the reciprocal isogenic F1 triploids, tiling array analysis of CG methylation in a CCGG context was performed. This analysis was performed using the same tiling array platform that was used for the transcriptome analysis, allowing for direct locus-by-locus comparisons to be made.

Briefly, DNA was extracted from matched samples (i.e. the four biological replicates used in the gene expression study) and the DNA digested with either of the DNA methylation-sensitive restriction enzymes *Msp*I or *Hpa*II, both of which recognize CCGG restriction enzyme sites. *Msp*I is insensitive to methylation at the internal cytosine (i.e. C^{m}CGG) and will cut the CCGG site whether the internal C sequence is methylated or not. In contrast, *Hpa*II will only cut CCGG when the internal cytosine is unmethylated (i.e. CCGG). Therefore, increased signal intensity at probes containing CCGG sequences in the *Hpa*II-digested samples indicates methylation at the internal C site (i.e. C^{m}CGG).

To scan the genome for CG methylation polymorphisms between the reciprocal F1 triploids, a total of 75 943 CCGG sites interrogated on the array were each tested for *Msp*I versus *Hpa*II intensity differences, and a linear, mixed-effects model was implemented to detect both constitutive and polymorphic methylation (see Materials and Methods, and ref.[37]). In this experiment, constitutive methylated CCGG sites are those in which internal cytosine methylation is observed in both of the isogenic reciprocal F1 triploids. In contrast, polymorphic methylated CCGG sites are those in which internal cytosine methylation is observed in only one of the isogenic reciprocal F1 triploids. Using this approach, 8008 CCGG sites were identified as methylated at the internal cytosine (i.e. C^{m}CGG) in both of the reciprocal F1 triploids (FDR $= 5.64 \times 10^{-3}$). In contrast, a further 5644 sites (7.4% of the total 75 943 sites scanned) displayed polymorphic

internal cytosine methylation, indicating that these internal cytosine sites were differentially methylated between the genetically identical F1 triploids. Of these, 3587 (63.6%) were methylated in the 3x(p) but not the 3x(m) (FDR = 3.54×10^{-3}), whereas the remaining 2057 (36.4%) showed the opposite pattern and were methylated in the 3x(m) (FDR = 5.77×10^{-3}) but not the 3x(p). Of the 13 652 C^{m}CGG sites detected across the genome, 41.3% were polymorphic between the reciprocal F1 triploids.

The extensive differential C^{m}CGG methylation between the isogenic reciprocal F1 triploids can be due to parental genome-dosage effects on ^{m}CG establishment or maintenance. Indeed, it is likely that there are parental genome-dosage effects on both of these DNA methylation mechanisms, as the 3587 could be neomethylated in the paternal-excess 3x, whereas the 2057 could be demethylated in the paternal excess (or vice versa). Our results suggested that RNA-directed DNA methylation (RdDM; via AGO4 and DRM2) at some loci could be subject to parent-of-origin-specific genome-dosage effects on CG methylation establishment.[38] However, the transcript levels of *AGO4* and *DRM2* do not differ between the reciprocal F1 triploids, indicating that the parent-of-origin-specific genome-dosage effects on ^{m}CG methylation polymorphism are not due to *AGO4*/*DRM2* transcript levels affecting ^{m}CG methylation establishment.

Our results could also indicate that maintenance of ^{m}CG methylation at some sites via *MET1*, *DDM1* and/or *VIM1-3* can be subject to parental genome-dosage effects.[38] However, the transcript levels of *MET1*, *DDM1* and/or *VIM1-3* do not differ between the reciprocal F1 triploids, indicating that the parent-of-origin-specific genome-dosage effects on ^{m}CG methylation polymorphism are not due to different transcript levels of *MET1*, *DDM1* and/or *VIM1-3* affecting ^{m}CG methylation maintenance. Alternately, the ^{m}CG methylation polymorphism observed could be due to parental genome-dosage effects on establishment or maintenance pathways that are operating at a post-transcriptional level. It is important to consider that the assay used does not interrogate all the CG sites in the genome, i.e. the assay is restricted to those CG sites in a CCGG context. Any conclusions made from our results hence relate to CCGG sites, and not to all CG sites in the genome. However, it is possible that the CCGG data subset may be representative of overall CG methylation (see below).

3.6. *Parental genome-dosage effect polymorphic C^{m}CGG methylation is uniformly distributed across chromosomal regions in reciprocal F1 triploids*

To determine whether the polymorphic ^{m}CG methylation at CCGG sites was clustered on any chromosomes or chromosomal regions, the genome-wide distribution of constitutive C^{m}CGG methylation was analysed across the five *A. thaliana* chromosomes. This determined that the constitutively methylated C^{m}CGG sites are concentrated in pericentromeric regions, while depleted in the middle of the chromosome arms (Fig. 2A). This pattern for constitutive C^{m}CGG methylation is similar to the distribution previously reported in diploids,[34,37] including assays evaluating ^{m}C in all contexts.[33] In contrast, the polymorphic C^{m}CGG methylation sites that are subject to parental genome-dosage effects (i.e. those which differ between 3x(m) and 3x(p) triploids) are much more uniformly distributed across the chromosomes (Fig. 2A).

As DNA methylation changes could have either site-specific or broader effects on local chromatin structure,[39] we further investigated whether the DNA methylation polymorphisms were clustered across large tracts of sequence rather than at individual sites (as our method identifies). Hence, LOWESS smoothing was performed on the *d* scores using 200 kb discrete windows for both constitutive and polymorphic C^{m}CGG methylated sites. LOWESS smoothing performs locally weighted regression on neighbouring *d* scores within the window such that each *d* score is adjusted to reflect the overall pattern displayed by its neighbours. This allows identification of regional methylation patterns along chromosomes when compared with a null distribution. After applying LOWESS smoothing, constitutive C^{m}CGG methylation continued to be preferentially located around the pericentromeres, indicating that constitutively methylated sites are clustered in these regions (Fig. 2B). In contrast, the LOWESS-smoothed *d* scores for polymorphic C^{m}CGG methylation sites are largely indistinguishable from those of the null distribution (Fig. 2C). This indicates that, in chromosome regions distal from the centromeres, polymorphic and constitutive C^{m}CGG methylation sites do not show any patterns specific to large chromosomal tracts (i.e. at 200 kb windows or multiples thereof), and that the CCGG sites, which are differentially methylated between F1 triploids, are therefore likely to be discretely located at a resolution of <200 kb.

3.7. *Polymorphic C^{m}CGG methylation is not associated with any particular genomic feature in the reciprocal F1 triploids*

It has been previously shown that $\sim$20% of genes are methylated in *A. thaliana* diploids.[34,35,37] We found that $\sim$20% of genes contained either constitutive or polymorphic C^{m}CGG methylation in the isogenic F1 triploids. To test whether the constitutive and polymorphic C^{m}CGG sites differed in their associations with other genomic features, we compared their distributions across coding sequences (CDS), introns, 5′ UTRs, 3′

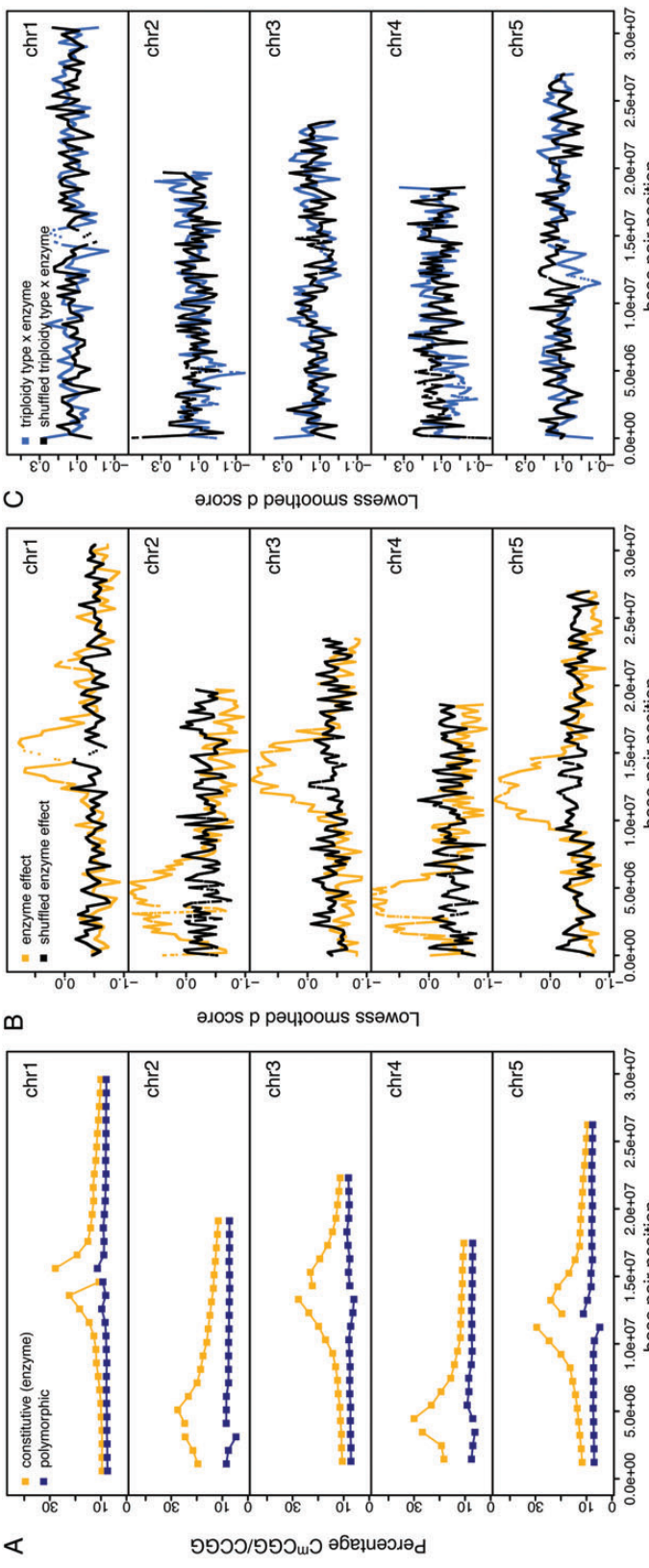

Figure 2. (A) Percentage methylation of constitutive (orange) and polymorphic (blue) methylation in a $C^{m}CGG$ context across the five chromosomes of *A. thaliana*. (B) LOWESS smoothing of constitutive methylation *d* scores using 200 kb discrete windows. LOWESS smoothing of constitutive methylation (orange), or a null distribution (black) obtained by shuffling by 1 kb blocks then LOWESS smoothing. (C) LOWESS smoothing of polymorphic methylation *d* scores using 200 kb discrete windows. LOWESS smoothing of polymorphic methylation (blue), or a null distribution (black) obtained by shuffling by 1 kb blocks then LOWESS smoothing. *X*-axis represents the position across the chromosomes.

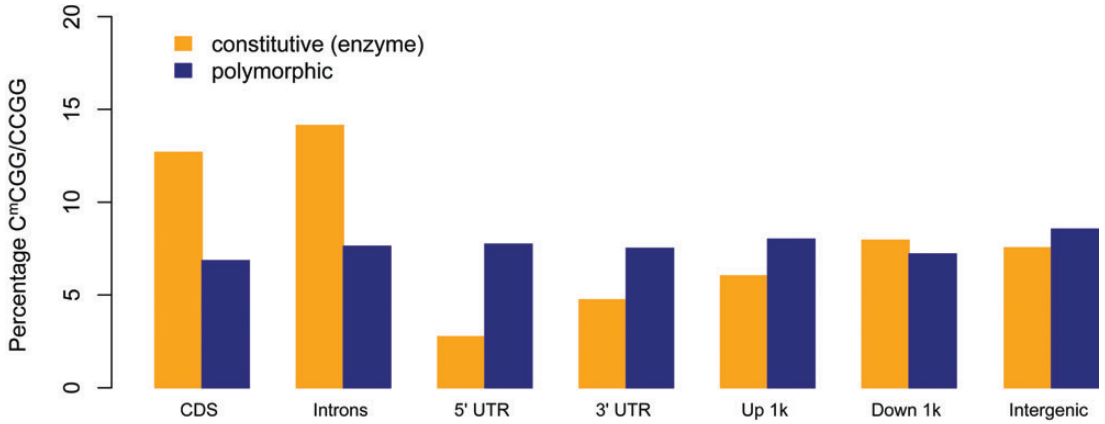

Figure 3. Percentage of methylated sites (C^{m}CGG) across genomic features for constitutive (orange) and polymorphic (blue) methylation. CDS: coding sequence; 5′ and 3′ UTRs: untranslated regions; up 1 kb: 1 kb region upstream of the genes; down 1 kb: 1 kb region downstream of the genes.

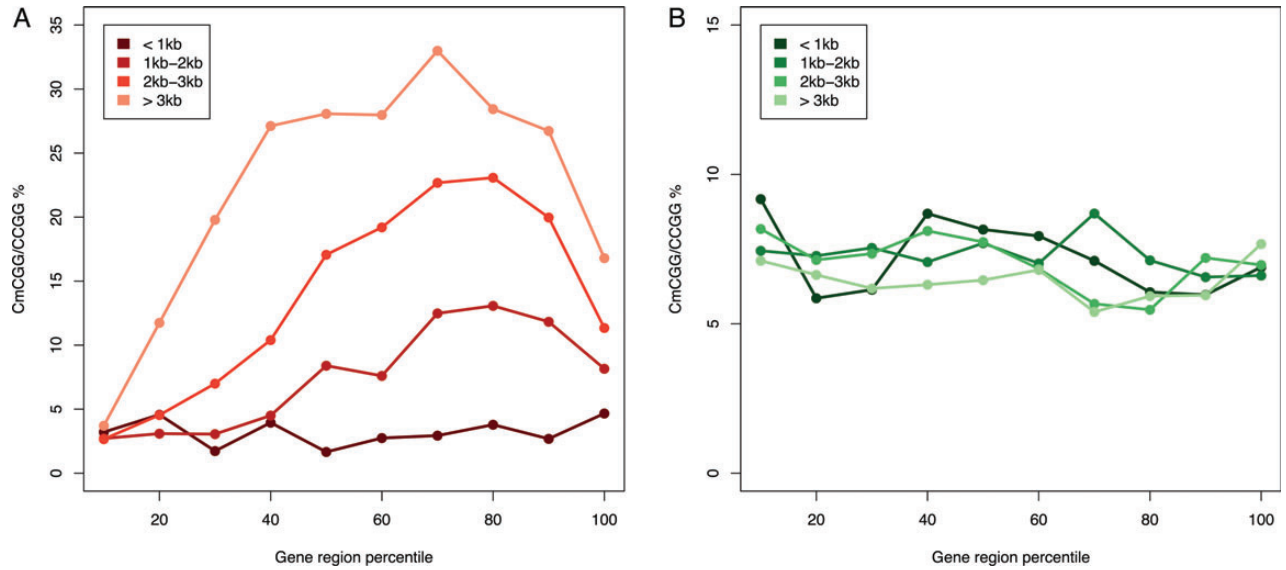

Figure 4. Percentage methylation (C^{m}CGG) across genic regions. (A) Constitutive methylation between reciprocal triploids and (B) polymorphic methylation between reciprocal triploids. Gene regions split into percentiles from 5′ to 3′ end of genes. Genes were further split into sizes ranges from <1, $1-2$, 2 and >3 kb.

UTRs, areas 1 kb upstream or downstream of genes, and in intergenic regions (Fig. 3). The constitutive C^{m}CGG methylation was found in all of the genomic features analysed, with some relative elevation in CDS and introns, and possible reduction in 5′ UTR and 3′ UTR (Fig. 3). The polymorphic C^{m}CGG methylation was found to a similar extent across all genomic features ($\sim$6–7%).

3.8. *Constitutive (but not polymorphic) C^{m}CGG gene body methylation in F1 triploids displays similar patterns to diploids*

While the role of gene body methylation has yet to be fully elucidated in plants, roles in transcriptional accuracy and splicing efficiency have been suggested,[34,40] and it appears that body-methylated genes are functionally important and display slower evolutionary rates in *A. thaliana*.[41] It has previously been shown in *A. thaliana* diploids that genes containing body methylation tend to be less methylated towards their 5′ and 3′ ends.[34,35,37] To determine whether this gene body methylation pattern was conserved in reciprocal F1 triploids, the distribution of both constitutive and polymorphic C^{m}CGG sites across the gene body (CDS and introns) was plotted (UTRs were omitted as they are not annotated for all genes). Genes were further divided by gene length (i.e. <1, $1-2$, $2-3$ and >3 kb) (Fig. 4A and B). Genes containing CCGG sites were divided into 10 percentiles,

and the proportion of C^{m}CGG methylation was calculated for each percentile.

The pattern of constitutive methylation observed in both reciprocal F1 triploids was similar to that observed in diploids,[34,35,37] with the 5′ and 3′ ends of body-methylated genes comparatively depleted in C^{m}CGG methylation. This pattern is particularly pronounced for longer genes (Fig. 4A), again confirming a pattern observed in diploids.[37] Furthermore, this pattern is observed when ^{m}C (i.e. methylation in all contexts) is plotted across gene bodies.[33] Given that gene body methylation in *A. thaliana* is primarily in a CG context,[42] this suggests that the CCGG sites tested in this study are broadly representative of gene body CG methylation in *A. thaliana*.

However, the distribution of polymorphic C^{m}CGG methylation across the gene body was rather different and was much more uniformly distributed across the gene body (Fig. 4B). There was also very little difference in the distribution of polymorphic C^{m}CGG methylated sites across genes of different lengths. In summary, total constitutive gene body methylation of gene body regions is higher than that of polymorphic methylation overall (Fig. 3) and is particularly higher in the middle of the gene body, and the 3′ end. In contrast, more C^{m}CGG polymorphisms are found at the 5′ end of genes when compared with constitutive methylation (Fig. 4A and B). We conclude that C^{m}CGG methylation patterns are largely unchanged in isogenic reciprocal F1 triploids generated in either cross direction (when compared with diploids).

A non-linear relationship between gene expression (absolute transcript levels) and gene body methylation has been observed in *A. thaliana*.[34,37] The least expressed genes and the most highly expressed genes are found to contain the lowest levels of gene body methylation, whereas genes expressed at intermediate levels contain the highest levels of gene body methylation.[33,36] To test whether this pattern of gene body methylation is maintained in *A. thaliana* F1 triploids, we investigated the correlation between constitutive C^{m}CGG methylation and absolute gene expression levels. All genes were divided into 20 percentiles categories according to their absolute expression levels (see Materials and methods). Within each expression level percentile and for a range of gene annotation categories (i.e. CDS, intron, 5′ UTR, 3′ UTR, up- and downstream 500 bp regions), the number of genes containing constitutive C^{m}CGG methylation site(s) was divided by those containing CCGG feature(s). For CDS and introns, the pattern of lowest and highest expressed genes having the lowest levels of constitutive C^{m}CGG methylation was observed (Fig. 5). In contrast, C^{m}CGG methylation within up- and downstream sequences, and in both UTRs, was generally low regardless of transcript expression levels (Fig. 5). As this

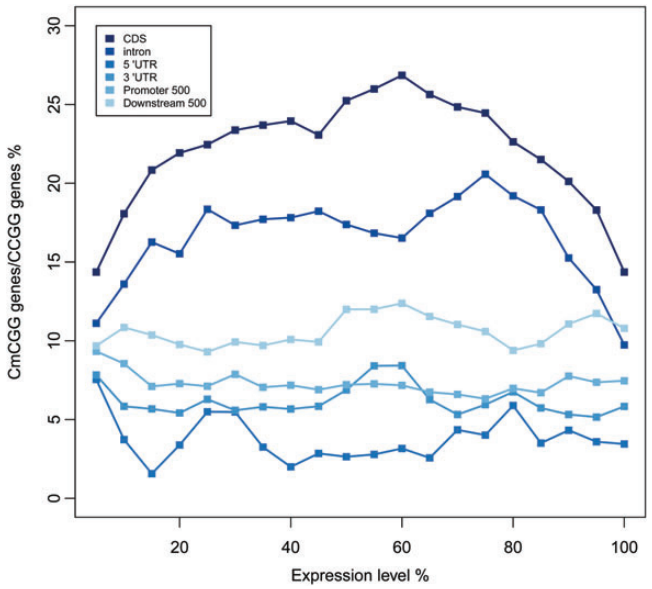

Figure 5. Association between constitutive methylation (C^{m}CGG) across genomic features (*y*-axis) and the absolute gene expression of associated genes (*x*-axis). Gene expression levels from all arrays [3x(m) and 3x(p)] were used and genes were split into percentiles.

pattern between gene expression and methylation is evident, when ^{m}C methylation is measured[34,35], it is a further indication that the CCGG sites measured in this study provide a good representation of the broader ^{m}CG methylation distribution. Although there remains the possibility that other important CG sites might be missed in analysis for particular loci.

3.9. *Gene expression differences between reciprocal F1 isogenic triploids are not correlated with differential C^{m}CGG methylation*

To examine the relationship between C^{m}CGG methylation polymorphisms and gene expression, gene expression *d* scores were linearly regressed against *d* scores for polymorphic methylation. Probes were mapped to genomic features CDS, introns, 5′ UTRs and 3′ UTRs, 1000 bp up- and downstream regions (up- and downstream regions were tested in 100 bp intervals). No significant correlation was observed between gene expression levels and polymorphic C^{m}CGG methylation at CDS, introns or downstream regions ($P < 0.05$). In contrast, a weak but significant negative correlation ($r = 0.13$, $P = 0.0039$) was identified between gene expression and regions 900–1000 bp upstream of genes.

Next, as most genes did not display any level of polymorphic methylation, we considered only those features found to have significant polymorphic methylation (either hypo- or hypermethylation), and compared their gene expression *d* scores with their polymorphic methylation *d* scores. Genomic regions hypomethylated in the 3x(p) were considered

separately from those regions hypermethylated in the 3x(p). A moderate positive correlation between hypomethylated regions 800–900 bp upstream of genes 3x(p) and gene expression was identified ($r = 0.31$, $P = 0.0459$). Overall, our results indicate that polymorphic C^mCGG methylation does not have a strong effect on gene expression in the reciprocal F1 triploids, and that the changes in gene regulation between paternal and maternal-excess triploids are not controlled by C^mCGG methylation.

3.10. *24-ht small RNAs differentially accumulate in up- and downstream regions*

Several species of small non-coding RNAs play a role in gene and transposable element regulation either by post-transcriptional gene silencing (miRNAs and tasiRNAs) or by RdDM (24-ht RNAs).[43] In the case of RdDM, the accumulation of 24-ht RNAs at a locus can be considered an indication of asymmetric CHH methylation at that locus (where H = C, A or T), as CHH methylation is reliant on RdDM. Using existing data of 3x(m) and 3x(p) small RNA sequencing from *A. thaliana* leaf tissue,[44] we investigated whether the differentially expressed genes between 3x(m) and 3x(p) displayed differential accumulation of either 21 or 24-ht RNAs across their genomic features (CDS, introns, 3′ UTRs, 5′ UTRs, up- and downstream regions).

For the genes we have as differentially expressed between the reciprocal F1 triploids, the number of accumulated 24-ht small RNAs was fewer in the upstream regions of dysregulated genes in the 3x(p) (median: 2.59, IQR: 1.29–61.10) compared with 3x(m) plants (median: 15.10, IQR: 1.89–91.52) (Fig. 6A), although the interquartile ranges are comparable, indicating that at least some upstream regions in the 3x(p) accumulate similar numbers of 24-ht RNAs compared with the upstream regions in the 3x(m) (Fig. 6A). Similarly, downstream regions of dysregulated genes were found to accumulate less 24-ht RNAs in the 3x(p) (median: 2.59, IQR: 1.29–78.23) compared with 3x(m) (median: 16.98, IQR: 1.89–111.80). In contrast, the other genic features tested displayed much smaller variations in 24-ht RNA accumulation (Fig. 6A). Furthermore, the accumulation of 24-ht RNA remained consistent across 3x(m) and 3x(p) for each feature tested (Fig. 6B).

To test whether these patterns of small RNA accumulation were specific for the differentially expressed genes

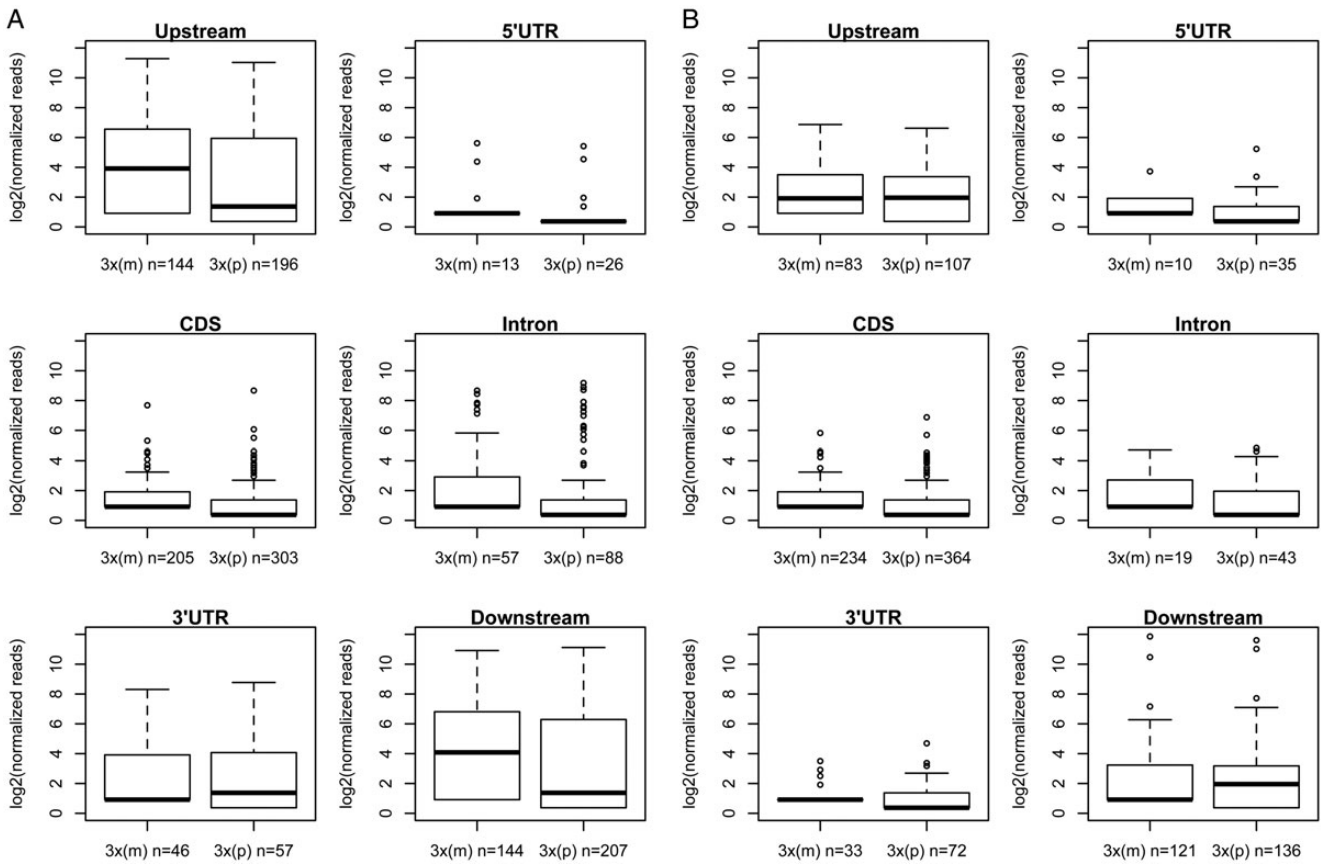

Figure 6. Boxplots of the distribution of small RNA accumulation across genomic features associated with genes differentially expressed between 3x(m) and 3x(p). n = the number of genes differentially expressed represented in each boxplot. (A) Boxplots of 24-ht small RNA accumulation. (B) Boxplots of 24-ht small RNA accumulation.

or whether they were consistent across all genes in the genome, the distribution of accumulation of 21- and 24-ht small RNAs for non-differentially expressed genes were identified (Supplementary Fig. S3A and B). The accumulation of both 21- and 24-ht RNAs was found to be comparable between 3x(m) and 3x(p) for all features tested, which is consistent with the original analysis of the small RNA data.[44] These results indicate that the loss of 24-ht RNAs in leaves in the up- and downstream regions is associated with differentially expressed genes up-regulated in the paternal-excess triploid. Indeed, differences in 24-ht accumulation in up- and downstream regions of genes have previously been shown to contribute to divergence in gene expression between two *A. thaliana* species.[45] Notably, 24-ht RNAs are known to be involved in *de novo* methylation, in particular asymmetric CHH methylation (via the RdDM pathway).[46] Hence, loss of 24-ht RNAs leading to loss of CHH methylation could be mechanistically involved in the up-regulation of loci gene expression in the paternal-excess F1 triploids. However, the small RNA dataset used here represents small RNA sequencing data from leaf tissues, which have comparative limitations. Further investigation on the distribution of small RNA in seedlings of reciprocal triploids will shed further light on the relationship between gene expression and small RNA distribution in triploids.

3.11. Conclusions

Overall, this study demonstrates that the transcriptomes of reciprocal F1 triploids are non-equivalent, despite the genetically identical nature of maternal genome excess versus paternal genome excess F1 triploids. This indicates that there are parent-of-origin-specific genome-dosage effects on the transcriptome of paternal-excess F1 triploids that could have an epigenetic basis. Even though DNA methylation is one form of epigenetic mark that has been widely associated with gene expression changes, our findings indicate that the parental genome-dosage-dependent effects on gene expression in paternal-excess F1 triploids are not associated with C^mCGG methylation and may instead be associated with RdDM pathways involving 24-ht small RNAs that are associated with *de novo* CHH methylation. Overall, our study suggests that the paternally and maternally inherited chromosome sets in autopolyploid plants may be epigenetically different, due to parental genome-dosage effects that can affect transcript levels in a C^mCGG methylation-independent manner. Such epigenetic differences between reciprocal F1 triploids that are genetically identical have implications for our understanding of triploidy and polyploidy in plants (and animals[47]), including for fundamental and applied genetics of triploid and other autopolyploid crops.

Supplementary Data: Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was funded through grant funding to C.S. from Science Foundation Ireland (SFI; grants 02/IN.1/B49 and 08/IN.1/B1931).

References

1. Birchler, J.A. and Veitia, R.A. 2012, Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines, *Proc. Natl. Acad. Sci. USA*, **109**, 14746–53.
2. Jiao, Y., Wickett, N.J., Ayyampalayam, S., et al. 2011, Ancestral polyploidy in seed plants and angiosperms, *Nature*, **473**, 97–100.
3. Chen, Z.J. 2010, Molecular mechanisms of polyploidy and hybrid vigor, *Trends Plant Sci.*, **15**, 57–71.
4. Madlung, A., Tyagi, A.P., Watson, B., et al. 2005, Genomic changes in synthetic *A. thaliana* polyploids, *Plant J.*, **41**, 221–30.
5. Verhoeven, K.J., Van Dijk, P.J. and Biere, A. 2010, Changes in genomic methylation patterns during the formation of triploid asexual dandelion lineages, *Mol. Ecol.*, **19**, 315–24.
6. Liu, B. and Wendel, J.F. 2003, Epigenetic phenomena and the evolution of plant allopolyploids, *Mol. Phylogenet. Evol.*, **29**, 365–79.
7. Buggs, R.J., Renny-Byfield, S., Chester, M., et al. 2012, Next-generation sequencing and genome evolution in allopolyploids, *Am. J. Bot.*, **99**, 372–82.
8. Riddle, N.C., Jiang, H., An, L., Doerge, R. and Birchler, J.A. 2010, Gene expression analysis at the intersection of ploidy and hybridity in maize, *Theor. Appl. Genet.*, **120**, 341–53.
9. Yao, H., Dogra Gray, A., Auger, D.L. and Birchler, J.A. 2013, Genomic dosage effects on heterosis in triploid maize, *Proc. Natl. Acad. Sci. USA*, **110**, 2665–9.
10. Adams, S., Vinkenoog, R., Spielman, M., Dickinson, H.G. and Scott, R.J. 2000, Parent-of-origin effects on seed development in *A. thaliana* require DNA methylation, *Development*, **127**, 2493–502.
11. Scott, R.J., Spielman, M., Bailey, J. and Dickinson, H.G. 1998, Parent-of-origin effects on seed development in *A. thaliana*, *Development*, **125**, 3329–41.
12. Miller, M., Zhang, C. and Chen, Z.J. 2012, Ploidy and hybridity effects on growth vigor and gene expression in *A. thaliana* hybrids and their parents, *G3 (Bethesda)*, **2**, 505–13.
13. Mittelsten Scheid, O., Afsar, K. and Paszkowski, J. 2003, Formation of stable epialleles and their paramutation-like interaction in tetraploid *A. thaliana*, *Nat. Genet.*, **34**, 450–4.
14. Guo, M., Davis, D. and Birchler, J.A. 1996, Dosage effects on gene expression in a maize ploidy series, *Genetics*, **142**, 1349–55.
15. Auger, D.L., Gray, A.D., Ream, T.S., Kato, A., Coe, E.H. Jr. and Birchler, J.A. 2005, Nonadditive gene expression in

diploid and triploid hybrids of maize, *Genetics*, **169**, 389–97.
16. Yu, Z., Haberer, G., Matthes, M., et al. 2010, Impact of natural genetic variation on the transcriptome of autotetraploid *A. thaliana*, *Proc. Natl. Acad. Sci. USA*, **107**, 17809–14.
17. Pignatta, D., Dilkes, B.P., Yoo, S.-Y., et al. 2010, Differential sensitivity of the *A. thaliana* transcriptome and enhancers to the effects of genome doubling, *New Phytol.*, **186**, 194–206.
18. Zhang, X., Yazaki, J., Sundaresan, A., et al. 2006, Genome-wide high-resolution mapping and functional analysis of DNA methylation in *A. thaliana*, *Cell*, **126**, 1189–201.
19. Kurihara, Y., Matsui, A., Kawashima, M., et al. 2008, Identification of the candidate genes regulated by RNA-directed DNA methylation in *A. thaliana*, *Biochem. Biophys. Res. Commun.*, **376**, 553–7.
20. Shen, H., He, H., Li, J., et al. 2012, Genome-wide analysis of DNA methylation and gene expression changes in two *A. thaliana* ecotypes and their reciprocal hybrids, *Plant Cell*, **24**, 875–92.
21. Boyes, D.C., Zayed, A.M., Ascenzi, R., et al. 2001, Growth stage-based phenotypic analysis of *A. thaliana*: a model for high throughput functional genomics in plants, *Plant Cell*, **13**, 1499–510.
22. Zhang, X. and Borevitz, J.O. 2009, Global analysis of allele-specific expression in *A. thaliana*, *Genetics*, **182**, 943–54.
23. Duszynska, D., McKeown, P.C., Juenger, T.E., Pietraszewska-Bogiel, A., Geelen, D. and Spillane, C. 2013, Gamete fertility and ovule number variation in selfed reciprocal F1 hybrid triploid plants are heritable and display epigenetic parent-of-origin effects, *New Phytol.*, **198**, 71–81.
24. Borevitz, J.O., Liang, D., Plouffe, D., et al. 2003, Large-scale identification of single-feature polymorphisms in complex genomes, *Genome Res.*, **13**, 513–23.
25. Zhang, X., Byrnes, J.K., Gal, T.S., Li, W.H. and Borevitz, J.O. 2008, Whole genome transcriptome polymorphisms in *A. thaliana*, *Genome Biol.*, **9**, R165.
26. Garnier, O., Laoueillé-Duprat, S. and Spillane, C. 2008, Genomic imprinting in plants, *Epigenetics*, **3**, 14–20.
27. Wolff, P., Weinhofer, I., Seguin, J., et al. 2011, High-resolution analysis of parent-of-origin allelic expression in the *A. thaliana* endosperm, *PLoS Genet.*, **7**, e1002126.
28. McKeown, P., Laouielle-Duprat, S., Prins, P., et al. 2011, Identification of imprinted genes subject to parent-of-origin specific expression in *A. thaliana* seeds, *BMC Plant Biol.*, **11**, 113.
29. Hsieh, T.F., Shin, J., Uzawa, R., et al. 2011, Regulation of imprinted gene expression in *A. thaliana* endosperm, *Proc. Natl. Acad. Sci.*, **108**, 1755.
30. Falcon, S. and Gentleman, R. 2007, Using GOstats to test gene lists for GO term association, *Bioinformatics*, **23**, 257–8.
31. Kim, E.D. and Chen, Z.J. 2011, Unstable transcripts in *A. thaliana* allotetraploids are associated with nonadditive gene expression in response to abiotic and biotic stresses, *PloS One*, **6**, e24251.
32. Wang, J., Tian, L., Lee, H.S., et al. 2006, Genomewide nonadditive gene regulation in *A. thaliana* allotetraploids, *Genetics*, **172**, 507–17.
33. Chao, D.Y., Dilkes, B., Luo, H., et al. 2013, Polyploids exhibit higher potassium uptake and salinity tolerance in *A. thaliana*, *Science*, **341**, 658–9.
34. Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. and Henikoff, S. 2007, Genome-wide analysis of *A. thaliana* DNA methylation uncovers an interdependence between methylation and transcription, *Nat. Genet.*, **39**, 61–9.
35. Vaughn, M.W., Tanurdzic, M., Lippman, Z., et al. 2007, Epigenetic natural variation in *A. thaliana*, *PLoS Biol.*, **5**, e174.
36. Cokus, S.J., Feng, S., Zhang, X., et al. 2008, Shotgun bisulphite sequencing of the *A. thaliana* genome reveals DNA methylation patterning, *Nature*, **452**, 215–9.
37. Zhang, X., Shiu, S.H., Cal, A. and Borevitz, J.O. 2008, Global analysis of genetic, epigenetic and transcriptional polymorphisms in *A. thaliana* using whole genome tiling arrays, *PLoS Genet.*, **4**, e1000032.
38. Dinh, T.T., O'Leary, M., Won, S.Y., et al. 2013, Generation of a luciferase-based reporter for CHH and CG DNA methylation in *A. thaliana*, *Silence*, **4**, 1.
39. 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 *A. thaliana* methylome, *Cell*, **152**, 352–64.
40. Regulski, M., Lu, Z., Kendall, J., et al. 2013, The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA, *Genome Res.*, **23**, 1651–62.
41. Takuno, S. and Gaut, B.S. 2012, Body-methylated genes in *A. thaliana* are functionally important and evolve slowly, *Mol. Biol. Evol.*, **29**, 219–27.
42. Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E. and Henikoff, S. 2005, DNA methylation profiling identifies CG methylation clusters in *A. thaliana* genes, *Curr. Biol.*, **15**, 154–9.
43. Chen, X. 2009, Small RNAs and their roles in plant development, *Annu. Rev. Cell. Dev. Biol.*, **25**, 21–44.
44. Lu, J., Zhang, C., Baulcombe, D.C. and Chen, Z.J. 2012, Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *A. thaliana* seeds, *Proc. Natl. Acad. Sci. USA*, **109**, 5529–34.
45. 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 *A. thaliana* and A. thaliana *lyrata*, *Proc. Natl. Acad. Sci. USA*, **108**, 2322–7.
46. Feng, S., Jacobsen, S.E. and Reik, W. 2010, Epigenetic reprogramming in plant and animal development, *Science*, **330**, 622–7.
47. Bogart, J.P. and Bi, K. 2013, Genetic and genomic interactions of animals with different ploidy levels, *Cytogenet. Genome Res.*, **140**, 1173.