siRNAs and DNA methylation: seedy epigenetics

Rebecca A. Mosher and Charles W. Melnyk

University of Cambridge Department of Plant Sciences, Downing Street, Cambridge, CB2 3EA, UK

To understand how DNA sequence is translated to phenotype we must understand the epigenetic features that regulate gene expression. Recent research illuminates the complex interactions between DNA methylation, small RNAs, silencing of transposable elements, and genomic imprinting in the Arabidopsis (Arabidopsis thaliana) seed. These studies suggest that transposable elements reactivated in specific cells of the gametophyte and seed might enhance silencing of transposable elements in the germline and embryo. By sacrificing genomic integrity these cells might make an epigenetic rather than genetic contribution to the progeny. This research could have implications for interspecies hybridization, the evolution of genomic imprinting, and epigenetic communication from plant to progeny.

An ever-expanding role for siRNAs and DNA methylation

With the completion of an increasing number of genome sequences, interest is growing in epigenetic phenomena: heritable phenotypes that are not explained by primary DNA sequence. Understanding how DNA modifications and chromatin structure affect gene expression is central to decoding genome sequence and has far-reaching implications for human disease and agriculture. Among the epigenetic phenomena under particular study are parental genomic imprinting (uniparental gene expression) and transcriptional gene silencing mediated by short interfering (si)RNAs, two phenomena which themselves are not widely understood. Recent papers describing DNA methylation and siRNA production in gametophytes and developing seeds of Arabidopsis provide clues linking these processes to silencing of transposable elements (TEs) [1–4]. One study showed that expression of RNA polymerase IV-dependent (p4-)siRNAs, many of which are generated from TEs, was maternal-specific in the endosperm [3]. Two further studies described coordinated genome-wide DNA demethylation of TEs in the endosperm and demonstrated that loss of CG methylation in these regions was associated with increased non-CG methylation and imprinted expression of neighbouring genes [1,2]. Finally, another report showed reactivation of TEs in the paternal gametophyte and accumulation of siRNAs in sperm cells [4]. Together these studies suggest that DNA demethylation and activation of TEs in accessory cells of the gametophyte and seed reinforces silencing of TEs in the germ line and embryo. Here we review what is known regarding p4-siRNA production and DNA methylation in the gametes and the seed, and how these factors interact to influence imprinted gene expression and control TEs. We propose that by sacrificing genomic integrity, accessory cells make an epigenetic contribution to the next generation to protect progeny from mobilization of transposable elements.

Double fertilization

Angiosperms diverged from non-flowering seed plants 125 million years ago and rapidly spread across the planet [5]. A key feature distinguishing flowering plants from their seed-plant predecessors is the production of two fertilized tissues in the seed (Figure 1) [6]. This process, termed double fertilization, might have provided an evolutionary advantage over non-flowering plants, although selection for this trait is poorly understood. In angiosperms, the maternal gametophyte contains the haploid egg cell and the homodiploid central cell, while the paternal gametophyte (pollen grain) contains two haploid sperm cells within the cytoplasm of the vegetative cell. The vegetative cell delivers the sperm cells to the maternal gametophyte where they fertilize the egg cell and central cell to generate the diploid embryo and the triploid endosperm respectively. The vegetative cell is critical for fertilization and the endosperm promotes embryo development and germination, but these accessory cells do not contribute genetically to the next generation.

RNA polymerase IV-dependent siRNAs

In addition to an elaborate fertilization procedure, flowering plants are distinct from basal plants due to expanded and diversified small RNA pathways [7], including the evolution of two additional DNA-dependent RNA polymerases [8–12]. RNA polymerase IV and V (Pol IV and Pol V) are plant-specific homologs of RNA polymerase II with specialized functions in small RNA production [13–16]. Pol IV initiates synthesis of 24 nt siRNAs from thousands of genomic locations [17,18] while Pol V is believed to transcribe intergenic regions and TEs [19]. P4-siRNAs are loaded into effector complexes and bind transcripts generated by Pol V to direct non-CG methylation at TEs [20]. However, the biological role of p4-siRNAs is not obvious because the loss of p4-siRNAs does not reactivates most TEs, and p4-siRNAs can also function in targeting DNA demethylation [17,21]. Although most p4-siRNAs are produced from repetitive elements of the genome, they are also associated with unique regions, including several characterized imprinted loci, and hence might play a role in initiating or maintaining imprinted gene expression.

Corresponding author: Mosher, R.A. (rm536@cam.ac.uk).
Parental genomic imprinting can refer to many parent-of-origin-dependent phenomena but generally refers to expression of a gene primarily or solely from only one parent’s allele. The resulting functionally haploid state can uncover deleterious mutations, making genomic imprinting a rare phenomenon previously confirmed at only six loci in Arabidopsis [22–27]. Parental genomic imprinting is restricted to angiosperms and therian mammals—organisms with a ‘placental habit’[28]. Just as the endosperm is a product of fertilization that does not contribute to the adult organism, the mammalian placenta is composed of zygotically-derived cells that are not incorporated into the adult. Interestingly, all imprinted genes in Arabidopsis are expressed in the endosperm[28] and many imprinted genes in mammals are expressed in placenta (even if imprinted elsewhere) [29], suggesting that the evolution of the placental habit is a prerequisite for the evolution of genomic imprinting [30,31].

Imprinted genes are associated with regions of differential DNA methylation (DMRs) between parental alleles. At all characterized imprinted genes in plants the maternal allele is hypomethylated regardless of which allele is expressed [28], indicating that DNA methylation is not simply silencing one allele but rather acting as a mark of parental origin. However, demethylation of maternal alleles must not always be the rule as loss of DNA METHYLTRANSFERASE 1 (MET1) in Arabidopsis gametophytes causes reciprocal phenotypes in seeds depending on parent of origin [32,33]. Loss of MET1-dependent maternal methylation causes phenotypes associated with excess paternal genomes (large seeds) while loss of MET1-dependent paternal methylation causes maternal-excess phenotypes (small seeds).

In Arabidopsis DMRs are created by the action of DNA METHYLTRANSFERASE 1 (MET1) in the paternal gametophyte and DEMETER demethylase (DME) in the maternal gametophyte [23,25,34,35]. Maternal methylation is first reduced through transcriptional repression of MET1 during female gametogenesis [36]. Subsequent expression of DME in the central cell but not in the egg cell generates DMRs only in the endosperm and restricts imprinted gene expression to this tissue. The observation that DMRs are associated with p4-siRNAs [37] indicates that siRNAs might direct MET1 to methylate the paternal allele in sperm cells or direct DME to demethylate the maternal allele in the central cell. However, changes in imprinted gene expression are not observed in p4-siRNA mutants [3], leaving open the questions of how MET1 and DME1 are targeted to establish DMRs, and what role p4-siRNAs play in imprinted gene expression.

p4-siRNAs are imprinted

Investigating the expression of p4-siRNA, we observed that many p4-siRNAs were restricted to flowers and immature seeds, where p4-siRNAs accumulated specifically in the endosperm [3]. Surprisingly, endosperm expression was completely eliminated when synthesis of p4-siRNAs in the maternal gametophyte was hindered, leading us to suspect that p4-siRNAs in the central cell directed chromatin modifications which initiated expression of p4-siRNAs in the endosperm. Because only maternal chromosomes are present in the central cell before fertilization, we predicted that expression of p4-siRNAs in the endosperm is maternal-specific. We confirmed this hypothesis by sequencing small RNAs from reciprocal crosses between two Arabidopsis ecotypes and detecting maternal-specific accumulation of p4-siRNAs from thousands of genomic locations. Because
levels of p4-siRNAs are much higher after fertilization than before, maternal-specific accumulation of p4-siRNAs cannot be a result of maternal-loading of the central cell, but must instead represent maternal-specific expression of p4-siRNA loci. The paternal genome is known to remain silent for 2–3 days after fertilization [38], but p4-siRNA expression remained maternal-specific at five days post-fertilization, indicating true genomic imprinting. The biological role of this high level of genomic imprinting is unknown but these results suggest that rather than causing imprinted gene expression, p4-siRNAs might be controlled by the same imprinting machinery as protein-coding genes.

**DNA demethylation in the endosperm**

Further support for the idea that p4-siRNAs might be regulated in a similar manner to imprinting genes comes from recent studies describing DNA methylation differences between the embryo and endosperm. Two groups profiled the global DNA methylation patterns in the embryo and endosperm of *Arabidopsis* to analyze epigenetic differences between the two products of fertilization [1,2]. One group sequenced methylcytosine immunoprecipitated DNA [1], while another group deep-sequenced bisulfite-converted DNA [2]. Both studies observed widespread reduction of DNA methylation across the endosperm genome compared to the embryo, particularly at TEs and regions generating p4-siRNAs. Both groups also demonstrated that loss of *DME* restored DNA methylation in the endosperm, although not to the high level observed in wild-type embryos. These data support the hypothesis that *DME* is responsible for the demethylation of TEs in the endosperm, but an additional pathway must operate to hypermethylate these sequences in the embryo.

To determine if any of the newly-identified DMRs cause imprinting of neighbouring genes, the authors of Ref. [1] searched for genes that are preferentially expressed in the endosperm and are near a region of *DME*-dependent hypomethylation. The authors verified imprinted expression for five such genes, nearly doubling the number of confirmed imprinted genes in *Arabidopsis* (Table 1). Interestingly, they found that only genes with endosperm-specific expression were imprinted, suggesting these genes might require demethylation via *DME* for expression. However, two of the newly identified imprinted genes display paternal-specific expression [1], suggesting that demethylation by *DME* suppresses the maternal allele in the endosperm. How maternal DNA demethylation causes locus-specific activation or repression is an interesting subject for future research (see Box 1). Extrapolating their results to the entire genome, the authors estimate that the total number of imprinted genes in *Arabidopsis* is approximately 50 [1], although this estimate may overlook imprinted genes with DMRs too small to be detected by methylcytosine immunoprecipitation, or imprinted genes with novel regulation.

**Dynamic DNA methylation**

Taking advantage of the single-base resolution of bisulfite sequencing, the context of DNA methylation across the endosperm and embryo genomes was analyzed [2]. The authors observed that TEs in the embryo were hypermethylated compared to vegetative tissues in both CG and non-CG contexts. TEs in the endosperm exhibited CG and non-CG methylation lower than the embryo, although non-CG methylation remained higher than vegetative tissues. In *dme* mutant endosperm CG methylation of TEs was restored to embryonic levels but surprisingly non-CG methylation was further reduced [2]. Because non-

**Box 1. Outstanding questions**

- How can DNA demethylation of the maternal allele result in both maternal-specific and paternal-specific expression at distinct loci?
- How is non-CG methylation affected in embryos and in the paternal endosperm genome of *dme* or *pol iv* mutants?
- What targets *DME* to TEs in the central cell? How does this relate to expression of p4-siRNAs?
- Do mobile siRNAs cause transcriptional or post-transcriptional silencing in sperm cells?
- Are these mechanisms conserved in other angiosperms?
CG methylation can be directed by p4-siRNAs, the authors postulate that DNA demethylation of the central cell by DME activates TEs, causing siRNA production and non-CG DNA methylation (Figure 2) [2]. Removing CG methylation to trigger non-CG methylation appears at first to serve little purpose. However, in addition to methylation of TEs in the endosperm, p4-siRNAs might move to the embryo to silence TEs, thereby triggering additional p4-siRNA production in the endosperm. It is also conceivable that an additional maternal-specific factor contributes to p4-siRNA expression in the central cell. A recent study has identified an RNA-binding protein that associates with ROS1, a homolog of DME [21], indicating the presence of RNA-directed DNA demethylation pathways in plants [17]. Future research into the targeting of DME and requirements for p4-siRNA production will undoubtedly resolve these questions (see Box 1).

Reactivated transposable elements in the paternal gametophyte

The central cell and endosperm are not the only terminal tissues with the potential to epigenetically regulate the next generation. In the paternal gametophyte the vegetative cell supports the sperm cells before fertilization. It was recently observed that TEs are reactivated and transposing in the Arabidopsis vegetative cell [4]. The reactivation of TEs was due to absence of the nucleosome remodelling protein DECREASED DNA METHYLATION1 (DDM1) from the nucleus of the vegetative cell, and at some TEs transcriptional reactivation was correlated with reduced DNA methylation and increased siRNA accumulation. The study further showed that small RNAs produced in the vegetative cell are capable of functioning in the sperm cells. The authors propose that loss of DDM1 in the vegetative cell causes upregulation of TEs and generation of siRNAs that move to sperm cells and enhance silencing of TEs in the germline (Figure 3) [4].

Whether these siRNAs are generated by Pol IV and whether they function in transcriptional or post-transcriptional gene silencing is unknown. The authors of Ref. [4] hypothesize that TEs are post-transcriptionally silenced in sperm cells due to the altered size of the siRNAs from Athila retrotransposons and loss of 24 nt siRNAs from other TEs. However, they observed strong non-CG methylation of TEs in sperm cells, a mark indicative of transcriptional gene silencing. A separate report found enhanced non-CG methylation of a class of retrotransposon in the vegetative nucleus and genetically implicated Pol IV and Pol V in this hypermethylation [40]. However,
Maternal influence over TEs in the seed

Although the paternal gametophyte might use a similar mechanism of sacrificing the genomic integrity of the vegetative cell to protect the sperm cells, it appears that epigenetic protection of the embryo is primarily a maternal responsibility. Because diploid maternal tissue supplies nutrients through the endosperm to the developing embryo, maternal investment in seed growth is higher than the paternal investment, favouring selection of maternal control of TEs in progeny.

It is intriguing to speculate that transport of small RNAs into the embryo also occurs in basal plants and gymnosperms. While non-flowering plants do not accumulate significant levels of 24-nt siRNAs [41–46], there is evidence for siRNA-mediated silencing of TEs in the moss Physcomitrella patens [47]. In mosses and non-flowering seed plants such as Ginkgo biloba, the sporophyte or embryo is nurtured by maternal gametophytic tissue and therefore any siRNAs transported into the embryo would necessarily be maternal. Maternal control over small RNAs in the seeds of angiosperms might mimic this evolutionarily ancient state. Alternatively the evolution of double fertilization might have allowed the paternal genome a conduit for epigenetic regulation of the embryo and the presence of siRNA transport in both the paternal gametophyte and the developing seed may reflect the contributions of both parents to the epigenetic state of the progeny. Investigating DNA methylation, siRNAs, and TE activity in a range of angiosperms and gymnosperms will ultimately shed light on these hypotheses (see Box 1).

**Interspecies hybridization**

The idea that the endosperm sacrifices genomic integrity to make an epigenetic contribution to the embryo raises the question of how such a mechanism functions in hybridization or wide crosses within the same species. Hybridization is particularly prevalent in angiosperms and many important crop species are hybrids [48] but the assimilation of divergent genomes can cause ‘genomic shock’ [49]. During Drosophila gametogenesis the nurse cells load piwi-associated (pi)RNAs into the egg to post-transcriptionally silence TEs from both maternal and paternal genomes after fertilization [50–53]. TEs contributed by the paternal genome that are not present in the maternal genome escape silencing and cause sterility by transposing in the hybrid’s germline. Reciprocal hybrids, in which the mother carries a TE not present in the paternal genome, develop normally. This effect is called hybrid dysgenesis and might be related to a process observed in hybrids between diploid Arabidopsis thaliana and Arabidopsis arenosa where activation of paternally-derived TEs is associated with failure of these hybrids [54]. Interestingly, hybrids are viable when maternal ploidy is increased, indicating that the paternal gametophyte contributes a dosage-sensitive regulator of viability. The role, if any, of Pol IV and p4-siRNAs in this process is unknown, but it is enticing to think that maternally-expressed p4-siRNAs are unable to silence divergent TEs from the paternal genome in the embryo or endosperm if contributed in low doses.

**Evolution of imprinted states**

It is known that integration of a TE near a gene can cause imprinting. This phenomenon was first observed in maize where the B-Bolivia allele of the b1 gene contains an upstream retrotransposon-like sequence and is imprinted in the endosperm, while the related B-Peru allele lacks the TE-like sequence and is not imprinted [55]. Similar allele-specific imprinting has been reported for other loci in maize [56–59] and it is reasonable to speculate that these alleles have also come under the control of a TE. Evolution of genomic imprinting after TE insertion has also been observed in mammals by comparative genetic analysis of eutherians, marsupials and monotremes [31,60–62]. However, to date researchers have failed to explain how integration of a TE causes imprinted expression. The recent work in Arabidopsis shows that TEs can be vectors for uniparental epigenetic modifications. By reactivating solely on maternal alleles in the endosperm, TEs in Arabidopsis can generate parental epigenetic asymmetry and imprinted expression when integrated near a protein-cod-
ing gene. It will be interesting to determine if TE silencing is also maternally-controlled in mammalian placentas.

The maize *maternally expressed in embryo 1* (*mee1*) is the first plant gene identified that displays imprinted expression in the embryo [63]. Interestingly, maternal-specific expression of *mee1* in the embryo is preceded by maternal-specific expression in the endosperm. Maternal and paternal *mee1* alleles in the embryo are equally methylated at fertilization and maternal alleles are actively demethylated in the embryo while maternal-specific expression occurs in the endosperm. The relationship between imprinted expression in the endosperm and embryo is unknown, as is the presence of p4-siRNAs at the *mee1* locus. Given the evidence for epigenetic communication within the seed, one might speculate that maternal p4-siRNAs move from the endosperm to embryo and recognize allele-specific chromatin marks to cause demethylation and transcriptional activation of maternal alleles. We have previously demonstrated similar chromatin-dependent RNA-directed DNA demethylation [17]. In this way, imprinting might evolve from being endosperm- or placenta-specific to occurring in somatic tissues.

**Epigenetic communication comes of age**

Recent research into the epigenetic states and small RNA populations of the male gametophyte and developing seed has uncovered new aspects of reproductive development. The vegetative cell and the endosperm might sacrifice genomic integrity to generate mobile small RNAs that influence TE silencing in the sperm cells or embryo [1–4]. Through this mechanism accessory cells could make an epigenetic contribution to the next generation. At the same time, epigenetic modification of maternal alleles before fertilization generates parental asymmetry after fertilization – a prerequisite for the evolution of imprinted expression.

The potential implications of this process are numerous. It is interesting to hypothesize that information gathered during an organism’s lifetime could be passed to the progeny using such a mechanism. Environmental stimuli might affect the siRNA population in accessory cells and establish the epigenetic state of the progeny, creating transgenerational ‘memory’ of the environment to which the parents were exposed [64,65]. Sacrifice of the vegetative cell and endosperm could also help re-establish ‘correct’ methylation patterns in the next generation [66]. There is also the possibility that mobile small RNAs have a role beyond reproductive development. RNA silencing in plants and animals can be mobile [67–69] suggesting that small RNA transport might exist in somatic tissues.

Future research integrating siRNAs, transcription, DNA methylation, and other epigenetic marks can help explain how cells communicate epigenetically and particularly how this communication has shaped the evolution of complex genomes and generated imprinted gene expression. Uncovering such an ‘imprinting code’ could allow mono-allelic expression of desirable genes in agriculturally important species such as hybrid corn, and translation of this research into mammalian systems might also generate insights to combat human imprinting diseases. As endosperm is the single largest contributor to human caloric intake, a better understanding of the function of this tissue will undoubtedly improve food stability.

**Acknowledgments**

The authors wish to thank David Baulcombe for his mentorship and support, and Donna Bond for comments on the manuscript. C.W.M. is funded by Commonwealth and NSERC scholarships.

**References**

conserved C-terminal domain of polyadenylate binding proteins. *Plant Cell* 20, 2387–2398
34 Choi, Y. et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell 110, 33–42
45 Dolgosheina, E.V. et al. (2008) Conifers have a unique small RNA silencing signature. RNA 14, 1508–1515
46 Motin, R.D. et al. (2008) Comparative analysis of the small RNA transcriptomes of *Poa contorta* and *Oryza sativa*. *Genome Res.* 18, 571–584
62 Pask, A.J. et al. (2009) Analysis of the platypus genome suggests a transposon origin for mammalian imprinting. *Genome Biol.* 10, R1