Endosperm and imprinting, inextricably linked

Mary Gehring\textsuperscript{1,2} * and P.R. Satyaki\textsuperscript{1}

\textsuperscript{1}Whitehead Institute for Biomedical Research and \textsuperscript{2}Massachusetts Institute of Technology, Cambridge MA

* corresponding author: mgehring@wi.mit.edu
ABSTRACT

The endosperm is often viewed as a complicated and rather strange tissue: its genome is triploid, it is formed from a second fertilization event specific to flowering plants, and it is a frequent source of incompatibility between plants, yet does not contribute genetic material to the next generation. Additionally, recent investigations have demonstrated that the endosperm is perhaps the most epigenetically divergent plant tissue, with unique DNA methylation and chromatin structure features. Endosperm is also the site of imprinted gene expression, an epigenetic phenomenon intimately linked to endosperm’s unusual properties. Imprinted genes are expressed predominantly from either the maternally or paternally inherited allele. This Update focuses on progress over the last several years in understanding imprinted gene expression, from molecular control to physiological significance, within the context of endosperm formation and function.
Seeds are a biological marvel – they house the next generation of plant and, in the right conditions, can remain viable for thousands of years (Sallon et al, 2008). The first seed plants appeared approximately 360 million years ago. About 200 million years later, flowering plants (angiosperms) arose and rapidly became the dominant form of plant life (Doyle, 2012). One of the key innovations in flowering plant seed development was the addition of a second fertilization event (Figure 1). While both flowering and non-flowering (gymnosperms) seed plants produce two sperm from each meiotic event, only one sperm participates in fertilization in most non-flowering seed plants. Resources are provisioned to the developing gymnosperm embryo through the female gametophyte, which extensively proliferates prior to fertilization (Figure 1). By contrast, in flowering plants the second sperm fertilizes a second gametic cell, the central cell, leading to formation of the endosperm (Drews and Koltunow, 2011). While many fascinating variations on female gametophyte development and the second fertilization event exist, in most flowering plants endosperm is triploid, consisting of two maternally inherited and one paternally inherited genome (Friedman et al, 2008). Although the idea remains unproven, the acquisition of a second fertilization event has been credited as one key to angiosperm dominance because the maternal parent devotes resources only to those seeds that have been successfully fertilized (Baroux et al, 2002).

The endosperm is the site of imprinted gene expression. In each plant species that has been examined in molecular detail, dozen to hundreds of genes that are either expressed from the maternally (maternally expressed imprinted genes or MEGs) or
paternally (paternally expressed imprinted genes or PEGs) inherited alleles have been identified (Figure 2). The embryo lacks imprinting (Gehring et al, 2011; Hsieh et al, 2011; Waters et al, 2011; Luo et al, 2011), except perhaps at the very earliest stages of development (Jahnke and Scholten, 2009; Nodine and Bartel 2012; Raissig et al, 2013), as do other plant tissues (Zhang and Borevitz, 2009). It is thought that imprinting is restricted to the endosperm because the formation of a sexual embryo-nourishing tissue may have created the right conditions for the evolution of genomic imprinting (Haig, 2013; Patten et al, 2014) (see Box 1). Major unanswered or only partially answered questions about imprinting include 1) how is imprinted expression controlled from the level of DNA sequence to three dimensional chromatin structure? 2) what is the function of imprinted genes? and 3) when and why did imprinted expression in endosperm evolve? Reflecting the current state of knowledge in the field, this Update focuses on the first two questions, with a smattering of the third, and primarily relies on insights gained from experiments performed in maize, rice, and Arabidopsis.

A BRIEF PRIMER on ENDOSPERM DEVELOPMENT and FUNCTION

Seed formation requires extensive intergenerational and intertissue communication and coordination (Ingram, 2010; Garcia et al, 2005; Ignouff et al, 2006; Li et al, 2013; Xu et al, 2016). Angiosperm seeds consist of three genetically distinct entities: the diploid embryo, with one paternal and one maternal genome, the triploid endosperm, with two maternal and one paternal genome, and the protective integuments/seed coat, consisting of diploid maternal tissue from the previous generation (Figure 1). After
fertilization, the primary endosperm nucleus (in the fertilized central cell) rapidly divides before the onset of zygotic division. In cereals, Arabidopsis, and many other species, initial endosperm divisions are synchronous and syncytial, meaning that there is nuclear division without concomitant cellularization. The surrounding maternal tissue also divides to accommodate the growing seed (Garcia et al, 2005; Li et al, 2013). After a defined period of nuclear division, endosperm cellularization in Arabidopsis coincides with reduced rates of mitotic division and the onset of endosperm differentiation (Mansfield and Briarty, 1990a; Mansfield and Briarty, 1990b; Boisnard-Lorig et al, 2001; Berger, 2003). Cellularization proceeds from the micropylar pole to the chalazal pole. Arabidopsis endosperm is distinguished into three tissue types: micropylar, which surrounds the embryo, peripheral, which lines the seed cavity, eventually displacing the large central vacuole, and chalazal, which abuts the termination of maternal vascular tissue (Mansfield and Briarty, 1990b; Boisnard-Lorig et al, 2001; Sorensen et al, 2002). Each endosperm tissue has its own gene expression program, with the chalazal endosperm being quite distinct from the micropylar and peripheral endosperm (Belmonte et al, 2013). The chalazal endosperm is characterized by larger nuclei, the result of nuclear fusions or endoreduplication (Boisnard-Lorig et al, 2001; Brown et al, 2003; Baroux et al, 2004), and is also the site of maternal nutrient transfer. In maize, endosperm is defined as ESR (embryo surrounding region, analogous to micropylar), starchy, aleurone, or BETL (basal endosperm transfer layer), which mediates nutrient transfer to the endosperm from the maternal parent (Sabelli and Larkins, 2009). Gene expression programs have also been defined for maize endosperm types at multiple
The endosperm is a food supply for the embryo/seedling (and other creatures). Recent exciting work has shown that the endosperm is not just a passive bag of nutrients, but is also a source of molecules that are required for proper embryo patterning. For example, the Arabidopsis and maize endosperm-specific basic-helix-loop-helix ZHOUP transcription factors are expressed in the ESR and are required for two genetically separable functions: embryo epidermis patterning and endosperm breakdown (Xing et al, 2013; Grimault et al, 2015). Short cysteine-rich peptides initially expressed in the central cell before fertilization and in the endosperm ESR after fertilization are required...
for elongation of the embryonic suspensor, a structure that connects the embryo proper to a nutrient supply during early seed development (Costa et al, 2014). Additional functions for endosperm and more cell non-autonomous acting molecules will likely be discovered over the next several years.

**IMPRINTING MECHANISMS and the UNIQUE ENDOSPERM EPIGENOMIC LANDSCAPE**

Imprinted genes offer a fascinating gene regulation puzzle – how do alleles of a gene, perhaps with identical sequence, living together in the same nucleus, exist in two different transcriptional states (**Figure 2**)? One possibility is that epigenetic differences are established before fertilization in the male and female gametophytes, and are faithfully propagated in the endosperm after fertilization; another is that differences are established post-fertilization. Several lines of evidence indicate that the embryo and endosperm exist in distinct transcriptional and chromatin states immediately after fertilization, and that these states are established during specification of the egg cell and central cell in the female gametophyte (Gutierrez-Marcos et al, 2006; Pillot et al, 2010; Garcia-Aguilar et al, 2010; Baroux et al, 2007a). Because sperm cells are capable of fertilizing either the egg cell or central cell (Frank and Johnson, 2009), differences between the zygote and the endosperm and between maternal and paternal DNA in the endosperm are likely established in the mature female gametophyte before fertilization. This is a fairly narrow time frame to establish dimorphism; the nuclei that will contribute
to the egg and central cell are not distinguished until the last mitotic division of the female gametophyte and occupy a common cytoplasm until cellularization.

Endosperm DNA is hypomethylated and has a looser chromatin structure compared to the embryo or other somatic tissues. Most observations of gametic and embryo and endosperm chromatin have been microscopic, and because of their inaccessible nature, there is a paucity of molecular details on the organization or epigenetic state of egg and central cell genomes. Development of cell or nuclei tagging approaches combined with decreasing amounts of material needed for chromatin, gene expression, or DNA methylation analysis is anticipated to rectify this within the next few years. At present, it is known that the central cell has low levels of H3K9me2, a histone modification usually associated with transcriptional repression, and lacks cytological chromocenters, in contrast to the egg cell or other somatic nuclei (Pillot et al 2010, Garcia-Aguilar et al, 2010). The central cell also lacks a histone variant, H2A.W, required for heterochromatin formation (Yelagandula et al, 2014).

DNA methylation is generally associated with transcriptional silencing, although many exceptions to this exist, particularly in endosperm (Makarevich et al, 2008; Gehring et al, 2009). The central cell is the site of DEMETER (DME) expression, which encodes a 5-methylcytosine DNA glycosylase that removes methylcytosine from DNA by base excision repair (Choi et al, 2002; Gehring et al, 2006). DME is an essential gene, and seeds that inherit a mutant maternal dm e allele undergo a prolonged period of
endosperm division and then abort about a week after fertilization (Choi et al, 2002). Because base excision repair disrupts nucleosomes, DNA demethylation probably both directly and indirectly leads to changes in chromatin modifications (Gehring et al, 2009b), although this has not been shown. The mechanism of targeting 5-meC DNA glycosylases to particular sites in the genome remains unknown.

The expression of the DNA methylation machinery is also low in the central cell (Jullien et al, 2012). In somatic tissues, DNA methylation is deposited by multiple pathways and is positively reinforced by chromatin modifications (Law and Jacobsen, 2010). The de novo DNA methyltransferases DRM2 and CMT2, which function in different but overlapping genomic contexts, establish methylation (Stroud et al 2013; Zemach et al 2013; Stroud et al 2014). Most methylation in plant genomes is found in transposable elements (TEs) or other repetitive DNA. DRM2 is guided to its targets by 24 nucleotide small RNAs, in a process termed RNA-directed DNA methylation (RdDM) (Matzke and Mosher, 2014). Somewhat counter intuitively, RdDM requires accessible chromatin in order to act, explaining its activity in more euchromatic regions (Schoft et al, 2009; Gent et al, 2014). DNA methylation often co-occurs with H3K9me2, and together these are important for repressing transposable element transcription. DNA methylation is maintained by the maintenance methyltransferase enzyme MET1 (CG methylation) and by the chromomethyltransferase CMT3 (CHG methylation). CMT3 interacts with H3K9me2, leading to positive reinforcement between histone methylation and DNA methylation. Asymmetric methylation (CHH) must be constantly established de novo
after DNA replication (Law and Jacobsen, 2010). Reporter gene fusions show that MET1 is not expressed in the central cell, and DRM2 expression is detectable but low (Jullien et al, 2012).

Epigenetic differences established in the central cell persist in the endosperm after fertilization. In the peripheral endosperm of young Arabidopsis seeds, an endosperm specific form of chromatin has been described, characterized by an absence of chromocenters and interspersed, diminished regions of heterochromatin (Baroux et al, 2007b). The decondensed chromatin appears to be driven by the maternal endosperm genome (Baroux et al, 2007b). Decondensed heterochromatin is associated with increased RdDM activity in other tissues (Schoft et al, 2009). Thus looser chromatin structure in the endosperm might be responsible for the abundance of small RNAs observed in seeds (Mosher et al, 2009). Most studies on DNA methylation in seeds have been performed between 6-8 days after pollination. At this stage, wild type endosperm is hypomethylated at TEs and related sequences, particularly at short TEs (i.e. fragments or remnants of TEs) that reside near genes (Gehring et al, 2009; Hsieh et al, 2009; Ibarra et al, 2012). Hypomethylation is specific to maternal alleles and depends on DME (Gehring et al, 2006; Ibarra et al, 2012). Thus, expression of DME in the central cell creates epigenetic asymmetry between maternal alleles in the embryo and endosperm, and between maternal and paternal alleles in endosperm (Gehring et al, 2006; Ibarra et al, 2012). Interestingly, endosperm from Arabidopsis dme mutants exhibits a complex methylation phenotype. Regions proximal to genes and/or
corresponding to short transposable elements are hypermethylated, as expected, but longer transposable elements are hypomethylated, contrary to expectations for an enzyme that removes 5-methylcytosine. These may be indirect effects, but suggest a requirement for DNA demethylation activity to maintain methylation in heterochromatin (Ibarra et al, 2012). Maize and rice endosperm DNA is also hypomethylated, and it is presumed that a DME homolog is responsible (Lauria et al, 2004; Waters et al, 2011; Zhang et al, 2014; Zemach et al, 2010).

By tagging and isolating nuclei of specific cell types in interspecific crosses, Moreno-Romero et al (2016) were recently able to examine allele-specific DNA methylation and histone modifications in Arabidopsis thaliana seeds much earlier than previous studies, at 4 days after pollination. Their findings suggest that DNA methylation is dynamic during endosperm development. CHH methylation is absent from endosperm at 4 DAP (Moreno-Romero et al, 2016). This agrees with gene expression profiling of laser capture dissected endosperm tissue that indicates low expression of the RdDM components in the endosperm until the heart stage (Belmonte et al, 2013), and with observations on the expression of methyltransferase reporter genes (Jullien et al, 2012). By 6 days after pollination CHH methylation is well established in the endosperm (Pignatta et al, 2014), indicating an increase in de novo DNA methylation activity.

Moreno-Romero et al (2016) also demonstrated that the maternal allele specific histone H3 trimethylation at lysine 27 (H3K27me3) is located at sites demethylated by DME. H3K27me3 is established by PRC2 (Polycomb repressive complex 2) and represses the
expression of developmental stage specific genes (Holec and Berger, 2012). PRC2 components are essential genes required for endosperm development, and prc2 mutants share some aspects of endosperm developmental (Choi et al, 2002) and DNA methylation (Ibarra et al, 2012) defects with dme. These data provide further impetus to fully describe and understand epigenome dynamics from gamete specification to seed maturity.

How do these unique endosperm features relate to gene imprinting? As of yet there is no clear understanding of how the cytological observation of decondensed endosperm chromatin relate to endosperm gene expression or imprinting, although it is thought that decondensation primarily affects the maternal genome (Baroux et al, 2007b). Techniques that would assay overall chromatin structure, like variations of chromosome conformation capture (Denker and deLaat, 2016), have not yet been applied to endosperm because of the large amounts of material required for these experiments. The relationship of DNA methylation to gene imprinting is better understood. Imprinted genes are often associated with differentially methylated regions (DMRs) between the maternal and paternal alleles, although there are many more DMRs in endosperm than there are imprinted genes (Gehring et al, 2009). Some MEGs have hypermethylated promoters in somatic tissues, like the Arabidopsis genes FWA and SDC, and demethylation of the maternal allele by DME promotes transcription in the central cell and endosperm. Silencing of the paternal allele depends on MET1 and RdDM (Kinoshita et al, 2004; Vu et al, 2013). However, this relatively simple regulatory
scenario describes a minority of imprinted genes. DNA hypomethylation also appears to influence transcriptional start site choice, leading to the production of different transcripts from maternal and paternal alleles (Luo et al, 2001; Du et al, 2014). The association between DMRs and imprinted genes in maize, rice, and Arabidopsis is much stronger for PEGs than for MEGs. Maternal alleles of PEGs are hypomethylated (Zhang et al, 2014; Rodrigues et al 2013, Pignatta et al, 2014). This was initially surprising – why is the hypomethylated allele the silent allele? Why is methylation required for expression of the active allele? It is now clear that the hypomethylated maternal alleles of most PEGs is also preferentially modified by H3K27me3 (Zhang et al, 2014; Du et al 2014; Moreno-Romero et al, 2016), which is generally correlated with repression. This suggests that DNA methylation might “protect” the paternal allele from gaining H3K27me3 (Weinhofer et al, 2010). In line with these observations, mutations in PRC2 components cause biallelic expression of PEGs (Hsieh et al, 2011). Interestingly, the epigenetic profile of PEGs is altered in A. lyrata, an outcrossing species closely related to A. thaliana (Klosinska et al, 2016). This species is also characterized by maternal genome hypomethylation, and many PEGs are associated with 5’ or 3’ DMRs. However, unlike A. thaliana, the silent maternal alleles of PEGs become CHG hypermethylated in the gene body. This phenotype has not been observed in wild type A. thaliana endosperm, but suggests that imprinting mechanisms may be flexible on relatively short evolutionary time scales.
Imprinting can be variable (also referred to as allele-specific) within species (Kermicle, 1978; Alleman and Doctor, 2000; Waters et al, 2013; Pignatta et al, 2014), and this may be related to variability in fidelity of epigenetic modifications. The sequences demethylated by DME and hypomethylated in Arabidopsis endosperm are enriched for transposable elements remnants, often of the Helitron and Mutator type (Gehring et al, 2009; Ibarra et al, 2012; Pignatta et al, 2014, Moreno-Romero et al, 2016). The presence of TEs or their methylation status can be variable within a species (Cao et al, 2011; Schmitz et al, 2013), which has implications for imprinting stability. Pignatta et al (2014) examined intraspecific variation in imprinting and found several examples of genes imprinted in one strain of *A. thaliana* but not another, which was correlated with variation in methylation of a nearby TE. These data suggest that gene expression programs in the endosperm might be particularly susceptible to natural epigenetic variation (Schmitz et al, 2013), because it is in endosperm that methylation state is relevant to the expression of some genes (Pignatta et al, 2014).

**IDENTITY and FUNCTIONS of IMPRINTED GENES**

Transcriptomic studies have identified imprinted genes on a genome-wide scale in the cereals rice, maize, and sorghum and in the dicots *Arabidopsis thaliana, Arabidopsis lyrata, Capsella rubella,* and castor bean (Luo et al, 2011; Zhang et al, 2011; Waters et al, 2011; Xin et al, 2013; Zhang et al, 2016a, Hsieh et al, 2011; Wolff et al, 2011; Gehring et al, 2011; Klosinska et al, 2016; Hantorangan et al, 2016; Xu et al, 2014). Typically, F₁ endosperm from crosses between two different strains is isolated at
specific stages of seed development, mRNA-seq performed, and imprinted genes identified by a deviation from the expected 2:1 ratio of maternal to paternal transcripts. Maternal and paternal transcripts are distinguished by sequence polymorphisms. Although the species that have been examined are vastly different in terms of genome complexity, organization, and epigenetic modification, and the criteria to define imprinted genes vary somewhat between groups, common themes have emerged.

The list of imprinted genes identified in any given experiment is significantly impacted by sequencing depth and the availability of sequence polymorphisms between the two parents (if a gene does not have a polymorphism in the coding sequence, parent-of-origin specific transcript information cannot be collected). Potential contamination from maternal integument/seed coat tissue remains a concern. Nevertheless, in each species approximately 75-200 imprinted genes have been identified. Protein coding genes, non-coding RNAs, and transposable elements are all subject to imprinting. Different sets of genes can be imprinted at different stages of seed development (Xin et al., 2013). This is perhaps not surprising given that each stage of seed development is typified by its own gene expression program, and understanding the imprinted gene universe for a species will require analysis at multiple stages of seed development.

Importantly, not all imprinted genes are monoallelicly expressed, meaning that expression between the parental alleles is differential, but not binary. (One could instead strictly define imprinted genes as those that are monoallelic, although this would
mean there were hardly any imprinted genes, and would exclude some genes identified as imprinted before the onset of genomics). Partial imprinting could arise from two scenarios: a mixture of imprinted and non-imprinted expression among individual endosperm cells or tissue types, or partial imprinting within each cell. These scenarios have bearing on understanding the molecular mechanisms of imprinting establishment and maintenance and imprinting function. This question could be addressed in greater detail by RNA-FISH or single cell transcriptomic studies.

As noted above, imprinting can also be variable within species, meaning that only some alleles of a gene are MEGs or PEGs. The first imprinted gene described, the maize R gene, exhibits allele-specific imprinting (Alleman and Doctor, 2000). Imprinting studies in maize and Arabidopsis that employed reciprocal crosses between multiple inbred lines concluded that about 10% of imprinted genes are variably imprinted (Waters et al, 2013; Pignatta et al 2014). The functional significance of this variation remains unclear. Some of the imprinting variation could be caused by rapid turnover of imprinting machinery as a result of conflict (see Box 1). Or, variable imprinting could reflect intraspecific differences in endosperm gene expression programs, as the timing of various aspects of seed development can be variable even within species. Another possibility is that allele specific imprinting is the result of relaxed selection on the imprinted state of a gene, suggesting that imprinting of that gene is not functionally relevant per se. Additional experimentation and understanding of gene function will be required to resolve these disparate possibilities for individual genes.
Analysis for enriched functional categories of imprinted genes have returned fairly modest results (Gehring et al, 2011, Wolff et al, 2011; Waters et al, 2011; Pignatta et al, 2014). In Arabidopsis and maize, PEGs are somewhat enriched for putative chromatin modifying functions or epigenetic regulators (Gehring et al, 2011; Pignatta et al, 2014; Waters et al, 2013). This is particularly intriguing because regulators of imprinting may themselves become part of an imprinting network under a conflict model of imprinting (see Box 1) (Patten et al, 2016). Many of these PEGs are predicted to promote DNA methylation or repressive histone modifications, including putative histone methyltransferases, components of the RdDM pathway, and genes required to maintain CG methylation, including VIM genes and MET1 homologs (Gehring et al 2011; Hsieh et al, 2011; Klosinska et al, 2016). By contrast, the PRC2 complex genes MEA and FIS2, which are required to maintain imprinted expression during endosperm development, are MEGs. In addition to proteins, imprinting regulators can include cis-regulatory non-coding RNA (ncRNA) as well as trans-regulatory small RNA. In maize, four ncRNA, speculated to have regulatory functions, are maternally expressed from within protein-coding loci that are paternally expressed (Zhang et al, 2011). In rice, small RNAs of the type that participate in RdDM and have the potential to regulate imprinting, are themselves imprinted (Rodrigues et al, 2013). However, it should be noted that apart from the PRC2 encoded proteins, the ability of these other factors to regulate imprinting remains to be demonstrated. This should be a fruitful area of future research.
Whether imprinting is functionally important for the plant or an inconsequential side effect of a relaxed endosperm epigenome has been debated (Berger et al, 2012). Mutations in several Arabidopsis PEGs have no obvious effect on seed development (Wolff et al, 2015). Yet the list of imprinted genes with clear functions in endosperm is growing. The imprinted gene *Meg1* is required for normal transfer cell development in maize BETL; altering the dosage of *Meg1* alters kernel size (Gutierrez-Marcos et al, 2004; Costa et al, 2012). The first imprinted genes discovered in Arabidopsis, *MEA* and *FIS2*, prohibit endosperm proliferation and are essential for normal endosperm development, although they also have important functions in the female gametophyte (Grossniklaus et al, 1998; Chaudhury et al, 1997; Kiyouse et al, 1999). One finding that could point to imprinting being “important” or functional is if homologous genes or genes that function in the same pathways were commonly imprinted among species. Comparisons of imprinting between maize and Arabidopsis or rice and maize report limited, but statistically significant, overlap among imprinted genes (Waters et al, 2013). FIE, another member of the PRC2 complex, is imprinted in rice and maize (Danilevskaya et al, 2003; Luo et al, 2009). Other conserved imprinted genes included auxin biosynthesis genes and epigenetic regulators, like the *VIM* maintenance methylation genes. Interestingly, *VIM103* is positively correlated with increased seed weight in a long-term selection experiment in maize (Zhang et al, 2016b). Endosperm-specific *YUCCA* genes, required for auxin biosynthesis, are PEGs in rice, maize, Arabidopsis, and related species (Luo et al, 2011; Waters et al, 2011; Gehring et al
Auxin is required for normal endosperm proliferation in Arabidopsis (Figueiredo et al, 2015) and in maize YUC1 positively regulates endosperm proliferation; plants with mutations in the gene produce lighter kernels with reduced endosperm cell number (Bernardi et al, 2012).

Because the absolute number of instances of conserved or convergent imprinting is few, the perception is that imprinting is rapidly changing among plant species. In practice, comparing imprinting between species is not trivial. Genes identified as imprinted in one species might not be called imprinted in another species for several technical reasons. It may be challenging to identify orthologues between distantly related species. Or, genes of interest might not have a SNP in all species, preventing discrimination between maternal and paternal alleles. Genes that are lowly expressed, as are some imprinted genes, are less likely to be represented in RNA-seq libraries. Additionally, the developmental stages of the seeds being used to assay imprinting may not be comparable between species. Thus, it is important to only compare imprinting among genes that can be confidently assessed for imprinting in both species (Klosinska et al, 2016). Two recent studies have compared imprinting between the closely related species A. thaliana and Capsella rubella, and A. thaliana and A. lyrata, which diverged from one another around 10-15 MYA (Hatorangan et al, 2016; Klosinska et al, 2016).

While one study characterized imprinting as rapid, and the other characterized it as fairly conserved, in both species comparisons imprinting of around 50% of genes was shared. PEGs appear to be more conserved than MEGs. Gene-focused comparisons of
imprinting between species do not take into account the possibility that other genes working in the same pathway could be imprinted, if the function of the pathway, rather than the function of an individual gene, is under selection. For example, in *A. lyrata* and *A. thaliana*, different components of the RdDM pathway are PEGs, and in maize and rice a different PRC2 component is imprinted than in Arabidopsis. On the whole, based on both comparisons of imprinting between species and on known function of imprinted genes, there appears to be ample evidence that genes subject to imprinting are important for endosperm proliferation and seed development, as will be further detailed below.

**IMPRINTING and INTERSPECIES or INTERPLOIDY HYBRID SEED PHENOTYPES**

Perturbation of endosperm development is a major source of seed lethality between different species and between plants of the same species but different ploidies. Although both embryo and endosperm development are defective in hybrid seeds, in many cases F₁ embryos can be rescued, suggesting that defects in endosperm development are casual for seed death and apparent effects on embryo growth are secondary (Ishikawa et al, 2011; Rebernig et al, 2015). Historical and more recent molecular evidence suggests a possible role for imprinting, or perturbation of imprinted gene expression, in both interploid and interspecific hybrid incompatibility. A key observation linking imprinting to hybrid seed endosperm failure is that reciprocal endosperm phenotypes are generated depending on which species or ploidy is the male parent in the cross, and which is the female. This is termed a parent-of-origin effect
(POE) and is classically associated with imprinting in plants and animals. It is important to note that POEs can arise for reasons other than imprinting, including cytonuclear interactions and gametophytic effects, and additional experiments must be undertaken to exclude these other possibilities before fingerling imprinting as the culprit (Kermicle, 1970).

A key developmental decision point during seed development is the timing of endosperm cellularization (Hehenberger et al, 2012) (Figure 3). Early cellularization leads to reduced endosperm size, whereas later cellularization or the absence of cellularization leads to larger endosperm. Both conditions can be lethal. For example, *Capsella rubella* and *Capsella grandiflora* diverged from one another only 100,000 years ago. When *C. grandiflora* is the female in interspecific crosses, hybrid seed lethality is associated with lack of endosperm cellularization. Some viable seeds (60%) are produced in the reciprocal cross, but these are small due to early endosperm cellularization (Rebernig et al, 2015). *C. rubella* is a self-fertilizing species and *C. grandiflora* is an outcrossing species, which may be related to the lack of completely reciprocal phenotypes (Rebernig et al, 2015; Brandvain et al, 2005). Similar phenotypes have been observed in rice interspecific crosses (Ishikawa et al, 2011), although in these instances the resultant seeds are larger or smaller than seeds from intraspecific crosses, rather than aborted. Monkeyflower (*Mimulus*), a genus with many recently diverged members that have overlapping ranges, is emerging as an interesting system for the study of hybrid endosperm effects. Pollination and fertilization occur equally well
between *Mimulus guttatus* and *Mimulus nudatus* as in self-fertilizations, but the resultant interspecific hybrid seeds are flat or shriveled because of the absence of or delayed endosperm proliferation (O’Neal, 2016). Using a clever breeding strategy, Garner et al recently mapped QTL responsible for hybrid seed lethality between *M. guttatus* and *M. tilingii* (Garner et al., 2016). Although the underlying genes have yet to be identified, many of the QTL effects were dependent on the parent-of-origin, further supporting the idea that imprinted genes might be involved. Several new mutants with parent-of-origin effects on maize endosperm development, albeit with incomplete penetrance, have also recently been identified (Bai et al 2016) and it will be informative to determine the molecular identity and imprinting status of the underlying genes.

Seed lethality is also a frequent outcome of interploidy crosses within species. In maize, maternal and paternal genome must be in a 2:1 ratio in the endosperm for the production of viable seeds (Lin, 1984; Pennington, 2008), and altering that dosage disrupts the transition from endosperm mitotic division to endoreduplication (LeBlanc et al, 2002). Other species, like Arabidopsis, are somewhat more tolerant of deviation from the 2:1 ratio (Scott et al, 1998). However, even when interploidy crosses are not lethal, endosperm syncytial division is either prolonged or attenuated based on whether excess genomes are inherited from the male or female parent. Maternal genomic excess is associated with early cellularization and paternal genomic excess with delayed (or in cases were seeds abort, the absence of) cellularization (Figure 3) (Scott et al, 1998; Pennington et al, 2008). The resemblance of reciprocal endosperm
phenotypes produced in interploidy and interspecific crosses has led to the idea that these are functionally similar phenomenon, both possibly related to gene imprinting (Lin, 1984; Bushell et al, 2003; Lafon-Placette and Kohler, 2016). How might this be so? In the framework of the popular kinship theory of imprinting (Box 1), MEGs are predicted to attenuate endosperm proliferation while PEGs promote it, and this is somewhat born out by the known functions of imprinted genes (see section above). In maternal genomic excess the dosage of active alleles for MEGs increases while the relative dosage of PEGs decreases, with the reverse true in instances of paternal genomic excess. Additionally, different genes might be imprinted in different species, leading to an altered gene expression program in F1 endosperm depending on which species serves as which parent.

Most molecular characterization of the gene expression changes that accompany interspecific and/or interploidy crosses have been performed in Arabidopsis thaliana and related species. In crosses between diploid A. thaliana females and diploid A. arenosa males, endosperm overproliferates and 95% of seeds abort. Seed lethality is significantly reduced if the A. thaliana female is tetraploid (Josefsson et al, 2006), suggesting, in congruence with the endosperm balance number hypothesis (Johnston et al, 1980; Kinoshita, 2007), that A. arenosa males are functionally tetraploid in relation to A. thaliana females. Expression of several PEGs, including the AGAMOUS-LIKE (AGL) MADS box transcription factor PHERES1 (PHE1), is upregulated from the normally silent maternal allele in the lethal cross, and mutation of phe1 in A. thaliana
partially ameliorates the hybrid incompatibility phenotype (Josefsson et al, 2006; Burkart-Waco et al, 2015). Failure of endosperm cellularization has been associated with increased expression of AGL MADS box transcription factors, in particular AGL62 (Kang et al, 2008). Other AGL genes also contribute to seed lethality, some of which are imprinted (Shirzadi et al, 2011; Walia et al, 2009). Many of the same genes are misregulated in interploidy crosses within A. thaliana (Erilova et al, 2009; Tiwari et al, 2010), further highlighting the molecular similarities of interploidy and interspecific seed lethality. Expression of the AGL genes is regulated by the PRC2 complex (Hehenberger et al, 2012), which is active in the female gametophyte and developing seeds. Seeds that inherit mutations in mea and fis2 from the female parent phenocopy interploidy crosses with paternal genomic excess (Figure 3), undergoing seed abortion after an abnormally prolonged period of endosperm syncytial division. Consistent with this, the prc2 mutant phenotype is suppressed if the maternal parent contributes two extra genomes (producing triploid embryos and pentaploid endosperms), presumably by restoring effective, rather than actual, genomic balance of maternal to paternal genomes in the endosperm (Kradolfer et al, 2013a).

A causal relationship between specific imprinted genes and seed abortion induced by interploidy crosses has recently been genetically demonstrated (Kradolfer et al, 2013b; Wolff et al, 2015). Taking a genetic approach, Kradolfer et al (2013b) identified mutations that suppressed the seed abortion produced when wild type A. thaliana females are pollinated by unreduced 2n pollen (a cross that normally generates inviable
seeds with triploid embryo and tetraploid endosperm). Paternal inheritance of mutations in the *ADMETOS (ADM)* gene, a PEG, causes partial normalization of endosperm cellularization timing and *AGL* expression, suppressing seed abortion (Kradolfer et al, 2013b). Mutations in *ADM* have no effect on seed development in traditional diploid seeds. *ADM* encodes a putative molecular chaperone whose expression is upregulated in triploid seeds destined for death. Interestingly, the increase in *ADM* expression is all from the paternal allele – imprinting of *ADM* remains intact. *ADM* is also a target of the FIS-PRC2 complex, and mutations in *adm* weakly suppress the *mea* seed abortion phenotype (Kradolfer et al, 2013b). Paternally inherited mutations in three additional PEGs, the putative histone methyltransferase *SUVH7, PEG2*, and *PEG9*, also normalize endosperm cellularization timing and partially suppress triploid seed abortion (Wolff et al, 2015). Like *ADM, SUVH7* and *PEG2* imprinting is maintained in triploid seeds, but expression levels are higher. Together, these data have led to the suggestion that imprinted genes establish interploidy hybridization barriers, ultimately promoting speciation (Gutierrez-Marcos et al, 2003; Lafon-Placette and Kohler, 2016). Although it is clear that imprinted genes can impact this process, it is less clear that parent-of-origin specific expression is the relevant feature with regard to hybrid seed failure. In many instances it is not loss of imprinting – i.e. both alleles being transcriptionally active (biallelic) or both alleles being transcriptionally silent – that is the outcome of hybrid crosses. Rather, as documented for the paternal allele of *ADM*, it is increased expression of the allele that is active in non-hybrid crosses. It has been argued that these effects are more about gene dosage than the specific parent-of-origin
from which a gene is expressed (Dilkes and Comai, 2004; Birchler, 2014). However, because imprinting is itself a type of dosage effect and imprinted genes are predicted to be dosage sensitive (Patten et al, 2014) it is difficult to distinguish these possibilities. If imprinting is instead viewed as a process required for expression of a gene (going from 0 active copies to 1 (PEGs) or 2 (MEGs) active copies), then increased expression of the expressed allele of imprinted genes in triploid seeds could be viewed as an imprinting effect.

**CONCLUDING REMARKS**

The past five years has seen an immense expansion of the plant imprinting field. Imprinted genes have been identified genome-wide in multiple species, and the relative ease of these experiments ensures that more will follow. The exploration of seed epigenome dynamics from fertilization to seed maturity is in its infancy, as is the understanding of what those dynamics mean with regard to imprinted gene expression or more broadly. The recent findings that imprinted genes can cause post-zygotic lethality nicely complement work on hybrid incompatibility stretching back to the first half of the 20th century. Many questions (see Outstanding Questions) remain on the horizon for this field. Advancing understanding will depend on a range of approaches, from deploying the latest genomic technologies to access and assess parent-of-origin specific endosperm chromatin structure and determine the relationship to active DNA demethylation, to detailed molecular genetic studies of regulation and function of individual imprinted genes.
Box 1: What selects for imprinted gene expression?
The neutralist perspective argues that imprinting serves no function in the endosperm and is only a consequence of the derepressed epigenetic state of the endosperm (Berger et al. 2012). However, the imprinted state of several genes is conserved between species (Luo et al. 2011; Waters et al. 2013; Hatorangan et al. 2016; Klosinska et al. 2016). This argues that imprinting is indeed under selection. Some prominent selectionist arguments for the existence of imprinting are mentioned here (Patten et al, 2014). Under the Parental (or the kinship) conflict model, a polyandrous mother improves her reproductive fitness by distributing resources equally among her progeny (Haig 2013). This drive comes in conflict with paternal alleles, which promote offspring success at the expense of matrilineal half siblings. This conflict selects for imprinting. Under this model, PEGs are likely to increase resource allocation and MEGs are likely to decrease resource allocation from mother to offspring (Haig 2013). Several imprinted genes that encode metabolite transporters, nutrient storage proteins and cell cycle regulators fit this model (Waters et al. 2013; Pignatta et al. 2014). Genes with less direct connections to resource allocation might also be involved if they regulate the expression of many other genes. Parental conflict promotes arms races with outcomes including increasing the expression of imprinted genes or imprinting new genes. Consistent with this idea, the expression level of PEGs is increased in A. lyrata compared to A. thaliana (Klosinska et al, 2016); as an outcrosser A. lyrata presumably has more parental conflict. An additional scenario that may be conflict related is in the context of cyto-nuclear interactions. It has been proposed that loci encoding proteins that interact with the mitochondria or the chloroplast will be selected for maternal expression (Wolf, 2009). Imprinting has also been proposed to evolve through other evolutionary pathways in a conflict independent manner (Spencer and Clark 2014). Imprinting of some loci may prevent parthenogenesis by blocking development prior to fertilization (Spencer and Clark 2014). *FIS2* and *MEA*, which repress the maternal alleles of PEGs
and prevent fertilization independent development, fit this model (Chaudhury et al. 1997; Kiyosue et al. 1999; Hsieh et al. 2011), as the imprinted auxin biosynthesis genes that are their targets (Figueiredo et al, 2015). Imprinting at some loci may **regulate the expression** of dosage sensitive genes (Dilkes and Comai 2004). The varied evolutionary pathways described in the models above might have all contributed to the evolution of the current crop of imprinted loci. The key task then is to assess the potential contributions of each pathway. Some of these models have theoretical support but finding empirical evidence is complicated. Many imprinted genes have no known phenotypes in the endosperm (Wolff et al. 2015) and the absence of a detectable function hinders one from guessing the evolutionary pathway that led to its imprinting. Many imprinted genes are paralogs and their imprinted state may have been selected to provide an incremental advantage for the maternal or paternal side in the conflict. In such a scenario, a mild phenotype may not be easily observed in the optimal conditions of the lab.

**ADVANCES:**
- Genome-wide surveys have identified imprinted genes in multiple monocots and dicots, allowing comparisons of imprinting conservation.
- Putative chromatin regulators, PRC2 components, RdDM components, and auxin biosynthesis genes are imprinted in monocots and dicots, suggesting functional importance of these genes for seed development.
- Cell type-specific chromatin, DNA methylation, and gene expression profiling is revealing unique features of the endosperm epigenome.
- The regulation of paternally expressed imprinted genes is becoming clearer – hypomethylated maternal alleles are repressed by PRC2 and the methylated paternal allele is expressed. The regulation of MEGs appears less cohesive.
- Imprinted genes prevent interploidy hybridization, possibly indicating an important role for imprinted genes in speciation.

**OUTSTANDING QUESTIONS:**
- What are the functions of imprinted genes and why does imprinting evolve at various loci? Are they imprinted through neutral evolution or through selectionist? If selection, what kinds of selection pressures are most dominant?
- Are genes with conserved imprinted states the result of convergent evolution or was their imprinting constantly maintained?
- What are the exact mechanisms by which imprinted genes and genome dosage regulate endosperm cellularization timing?
What is the spatial nuclear organization of endosperm chromatin? How does this relate to gene imprinting? Is decondensed endosperm chromatin dependent on active DNA demethylation?

Which came first: central cell DNA demethylation or gene imprinting?

What is the mechanism of imprinting for genes with no obvious connection to allelic DNA methylation or histone methylation differences?

**Figure Legends**

**Figure 1:** Endosperm is an important angiosperm innovation. In gymnosperms, a haploid sperm fertilizes one of the haploid eggs. The embryo develops surrounded and nourished by female gametophyte tissue. In angiosperms, one haploid sperm fertilizes the egg and the other fertilizes the diploid central cell. The central cell is diploid because two haploid nuclei, the polar nuclei, migrate toward one and fuse either before (Arabidopsis) or at the time of (maize) fertilization. The diploid embryo is nourished by the triploid endosperm. A single layer of endosperm remains in mature Arabidopsis seeds; in maize the endosperm supports the germinating seedling.

**Figure 2:** Imprinted genes can be either MEGs or PEGs. Imprinted genes are expressed in a parent-of-origin dependent manner. Although the diagram depicts complete silencing of one allele, in practice many imprinted genes are differentially expressed between maternal and paternal alleles, not monoallelically expressed.

**Figure 3:** Timing of endosperm cellularization is a key developmental decision point. Interspecies and interploidy crosses alter when or if endosperm cellularizes, which can lead to smaller or larger seeds, or even seed abortion. Mutations in some epigenetic regulators can partially ameliorate cellularization defects.

**Acknowledgements**

This work was supported by National Science Foundation grant MCB 1453459 to M.G.

**References**


Mansfield SG and Briarty LG (1990a) Development of the free-nuclear endosperm in *Arabidopsis thaliana* (L.) *Arabidopsis Info. Serv.* **27**: 53-64.


O’neal E, Willis JH, Franks RG (2016). Disruption of endosperm development is a major cause of hybrid seed inviability between *Mimulus guttatus* and *Mimulus nudatus*. *New Phytol.* **210**: 1107-1120.


Figure 1: Endosperm is an important angiosperm innovation. In gymnosperms, one of the haploid eggs is fertilized by a haploid sperm. The embryo develops surrounded and nourished by female gametophyte tissue. In angiosperms, one haploid sperm fertilizes the egg and the other fertilizes the diploid central cell. The central cell is diploid because two haploid nuclei, the polar nuclei, migrate toward one and fuse either before (Arabidopsis) or at the time of (maize) fertilization. The diploid embryo is nourished by the triploid endosperm. A single layer of endosperm remains in mature Arabidopsis seeds; in maize the endosperm supports the germinating seedling.
**Figure 2:** Imprinted genes can be either MEGs or PEGs. Imprinted genes are expressed in a parent-of-origin dependent manner. Although the diagram depicts complete silencing of one allele, in practice many imprinted genes are differentially expressed between maternal and paternal alleles, not monoallelically expressed.
**Figure 3: Timing of endosperm cellularization is a key developmental decision point.**

Interspecies and interploidy crosses alter when or if endosperm cellularizes, which can lead to smaller or larger seeds, or even seed abortion. Mutations in some epigenetic regulators can partially ameliorate cellularization defects.