Epigenetic mechanisms regulating genomic imprinting in rice seeds

by

Jessica Astrid Rodrigues

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Plant Biology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Associate Professor Daniel Zilberman, Chair
Professor Robert Fischer
Professor Michael Freeling
Professor Jasper Rine

Fall 2014

Epigenetic mechanisms regulating genomic imprinting in rice seeds

by

Jessica Astrid Rodrigues

Abstract

Epigenetic mechanisms regulating genomic imprinting in rice seeds

by

Jessica Astrid Rodrigues

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Associate Professor Daniel Zilberman, Chair

Genomic imprinting is a phenomenon that has evolved in certain sexually-reproducing clades, including mammals and flowering plants. It is characterized by parent-of-origin-dependent expression of certain loci due to differing chromatin signatures or 'imprints' between maternally-inherited and paternally-inherited DNA. Previous work in the endosperm of plants such as maize and *Arabidopsis thaliana* showed that maternal DNA hypomethylation serves as the primary imprint for imprinted expression at certain genes. Subsequent genome-wide surveys of DNA methylation in *Arabidopsis thaliana* endosperm uncovered extensive localized DNA demethylation on maternally-inherited but not paternally-inherited chromosomes. Though this demethylation mediates imprinted expression of both maternally- and paternally-biased genes, it is apparently unnecessary for the extensive accumulation of maternally-biased small RNA molecules detected in *Arabidopsis* seeds. Whether imprinting function and genome-wide patterns of imprinting regulation are conserved across monocots and dicots remains to be tested.

Here, I show that extensive localized hypomethylation of rice endosperm DNA is likewise due to hypomethylation of maternally-inherited but not paternally-inherited chromosomes. Maternal hypomethylation preferentially occurs at regions of high DNA accessibility and is enriched within both imprinted genes and imprinted small RNA-producing loci. Maternally expressed imprinted genes are enriched for hypomethylation at putative promoter regions and transcriptional start and end sites, while paternally expressed genes are enriched for hypomethylation at promoters and gene bodies, mirroring recent results in *A. thaliana*.

I also show that, unlike in other rice tissues, small RNA populations in rice endosperm are dominated by a limited number of strong small RNA-producing loci. Whereas small RNAs in rice seedling tissues primarily originate from small Class II (cut-and-paste) transposable elements, those in endosperm are more uniformly derived, and include sequences from other transposon classes, as well as genic and intergenic regions. Imprinted 24-nt small RNAs are expressed from either parental genome and correlate with maternal DNA hypomethylation. Overlaps between imprinted small RNA loci and imprinted genes expressed suggest that small RNAs are associated with parent-of-origin-specific silencing.

In order to further investigate imprinting mechanisms and their targets in rice, I analyzed imprinting divergence among four cultivars that span the diversity within the rice species *Oryza sativa*. While the imprinting of 395 out of 413 genes is conserved among rice cultivars, I estimate that 4 to 11% of imprinted genes show imprinting divergence. For 16 out of 20 genes with diverged imprinting, DNA methylation epialleles were observed in key regulatory regions identified by our genome-scale enrichment analysis. These regions included the promoter and transcription start site for maternally expressed imprinted genes, and the promoter and gene body for paternally expressed imprinted genes. I did not observe an obvious association of imprinted small RNAs with imprinting divergence. In general, DNA methylation and small RNA profiles are conserved among rice cultivars.

To test whether genetic mutations contributed either to the formation of epialleles associated with imprinting divergence or to imprinting divergence that could not be correlated with an epiallele, I assessed genetic variation at seven of the 20 loci with diverged imprinting. At three genes, imprinting divergence was due to the insertion or precise excision of retrotransposons. DNA methylation epialleles in key regulatory regions at the other four loci did not appear to be associated with genetic variation. I also attempted to evaluate whether some apparent variation in parental bias might be due to a resetting of the imprint leading to biallelic expression in later endosperm development, as has been suggested by another group. However, I could not correlate changes in parental bias with changes in DNA methylation and small RNA production, suggesting either that developmental resetting occurs through another mechanism, or that the analysis was confounded by a mixed population of endosperm cells, or that the apparent variation is a result of technical artefacts rather than imprint resetting.

In summary, I identified the key regulatory regions of maternally- and paternally-expressed imprinted genes where maternal DNA hypomethylation is associated with imprinting, and developed a model for imprinted expression where maternal DNA hypomethylation serves as a primary imprint at both maternally- and paternally-expressed loci. Recent data supports the applicability of this model to other plant species such as maize and *Arabidopsis*. Divergence in imprinting in rice is associated with disruptions to either the genetic sequence or epigenetic state of regulatory regions, with retrotransposons playing a major role in sequence-related imprinting divergence over short evolutionary scales. Further work is required to elucidate the specific biogenesis and function of imprinted small RNAs and to verify potential cases of imprint resetting. The findings I present here make contributions to current understanding and models of mechanisms of epigenetic regulation, seed development, hybrid incompatibilities, and plant evolution. The production of high quality seeds is essential to many agricultural ventures, both as a mechanism for propagating plants and as means of generating valuable food commodities. In combination with other research, my work has the potential to improve breeding strategies for cereal crops that feed much of the world.

Contents

List of Figures

List of Tables

Acknowledgements

This body of work would not have been possible without the genius, kindness, patience, hard work, and love of many people. A special thanks goes out to the cheerful team at Pam Ronald's lab at Davis who collaborated with me on this project, especially Randy Ruan, who performed the nightmarish task of crossing together various strains despite his severe allergies to rice pollen.

I owe a debt of gratitude to past and present members of the Zilberman and Fischer Labs who taught me about chromatin and seed biology, helped shape me as a scientist, and regularly helped me vent the frustrations of research. I am especially indebted to Assaf Zemach, Jason Huff, Ping-Hung Hsieh, Tzung-Fu Hsieh, Christian Ibarra, and last but not least, Daniel Zilberman and Bob Fischer. As iron sharpens iron, so one person sharpens another, and I am very fortunate to have experienced times of exciting advances in chromatin biology and gene regulation in these labs, and to have learned from the best.

The Department of Plant & Microbial Biology in general has been a great place to be a graduate student and I have greatly benefited from the friendships and intellectual stimulation contributed by my peers. Growing up in the Middle East and immigrating with my family to New Zealand, I never imagined that I would move half way across the globe (again) and feel as at home as I did. To my adopted family in Berkeley (Charlotte Carlstrom, Robert Rohde, and Stephen Yee): thank you for making this my home away from home. I also thank the Fulbright selection committees both in the U.S. and in New Zealand for this life-changing opportunity and all the resources provided to me as part of the program. My fellow Fulbrighters have been a truly inspirational bunch and I am honored to be part of that community.

I also extend my deepest appreciation to the various teachers and mentors I have had at different stages in my life, people who encouraged me to think critically and to strive to make a difference: Sunipa Neogi, Marian Brendish, Don Love, Shiva Reddy, Sushma Prakash, and Kathy Mountjoy.

Finally, I should acknowledge the people dearest to me, the people who cheered me on through my successes, picked me up during my falls, and contributed every effort towards helping me be the best I can be. To my parents, James and Maria Rodrigues: thank you for believing in me even before I could recite the alphabet and for being the best teachers and role models I could have asked for. Thank you to my 'baby sisters', Sophia and Samantha Rodrigues, whose sense of humour, love of learning, and calmness in times of trouble made them wonderful 'collaborators' for my earliest 'experiments'. And the last vote of thanks goes to my wonderful husband, Sunjay Galen: thank you for being my rock and for the selfless way in which you support me.

Chapter I

Introduction

The discovery of genomic imprinting

Genomic imprinting is a phenomenon seen in a subset of sexually-reproducing organisms. It is defined by biased expression of a gene from one parentally-inherited allele over the other, even though both alleles are present in the same cell and may even have identical DNA sequences. A gene that is predominantly expressed from the maternally-inherited allele is referred to as a maternally expressed imprinted gene (often abbreviated to MEG) while a gene that is expressed from the paternally-inherited allele is referred to as a paternally expressed imprinted gene (PEG) (H. Jiang & Kohler, 2012). The implication for both MEGs and PEGs is that maternally- and paternally-inherited alleles are differentiated post-fertilization by a mark that is epigenetic, or in other words, not directly related to DNA sequence. This epigenetic mark is referred to as the 'imprint' for the imprinted gene (H. Jiang & Kohler, 2012).

Evidence of genomic imprinting was observed long before modern concepts of genetics developed. Historical records from over 3000 years ago indicate that mule breeders in Asia Minor noted that different phenotypes were produced in the offspring of crosses between horses and donkeys depending on whether the maternal parent was a horse and the paternal parent was a donkey or vice versa (Morison & Reeve, 1998). In the wake of the rediscovery of Gregor Mendel's work in the early twentieth century (Mendel, 1901), evidence of genomic imprinting was observed as a case of non-Mendelian inheritance and grouped in a general category of effects due to different hereditary contributions of male and female gametes, termed anisogeny (Bateson, 1926). Work in the mealybug insect, *Pseudococcus nipae*, provided an example of different chromatin states of the same chromosomes depending on whether they were in the male gamete or female gamete, and was the first time the word "imprinting" was used to described such events (Schrader, 1921). However, imprinting was not described in plants and mammals till years later as it was mostly thought that all the differences in hereditary contributions of plant and animal male and female gametes were due to different genetic contributions.

In 1970, the concept of genomic imprinting in plants began to take shape, when Jerry L. Kermicle described a gene that causes mottled pigmentation in maize kernels only when inherited paternally (Kermicle, 1970). Kermicle showed that this gene was passed down through both male and female gametes, but was only able to effect a phenotype when inherited from the paternal parent. After Kermicle's discovery in maize, formal descriptions of imprinted genes were made for the first time in mammals with three landmark publications in 1991: descriptions of the selective maternal expression of the mouse Igf2r gene (Barlow, Stoger, Herrmann, Saito, & Schweifer, 1991), the paternal expression of the mouse Igf2 gene (DeChiara, Robertson, & Efstratiadis, 1991), and maternal expression of the mouse H19 gene (Bartolomei, Zemel, & Tilghman, 1991). Subsequent findings from genetic experiments in both plants and mammals (Huh, Bauer, Hsieh, & Fischer, 2008; Morison & Reeve, 1998) lead to a greater understanding of the mechanisms of imprinting and functions of imprinted genes, but it was not till the advent of high-throughput sequencing that comprehensive investigations of imprinting could be made.

Current evidence indicates that the phenomenon of imprinted expression of specific loci on chromosomes that are largely biallelically expressed is only observed in organisms with a 'placental habit': placental mammals and flowering plants (Pires & Grossniklaus, 2014). In mammals, the timing of imprinting evolution is thought to be linked to the spread of long terminal repeat (LTR) retransposons and the evolution of the placenta, as shown in Figure 1.1 (Renfree, Suzuki, & Kaneko-Ishino, 2013). Plant lineages have been less widely sampled for imprinting, but current limited evidence suggests that imprinting may have evolved at the same time as the endosperm (Pires & Grossniklaus, 2014), as shown in Figure 1.2. As imprinting in both cases is a result of processes specific to the male and female germlines and to the developing zygote, a brief overview of sexual reproduction in plants and mammals follows.

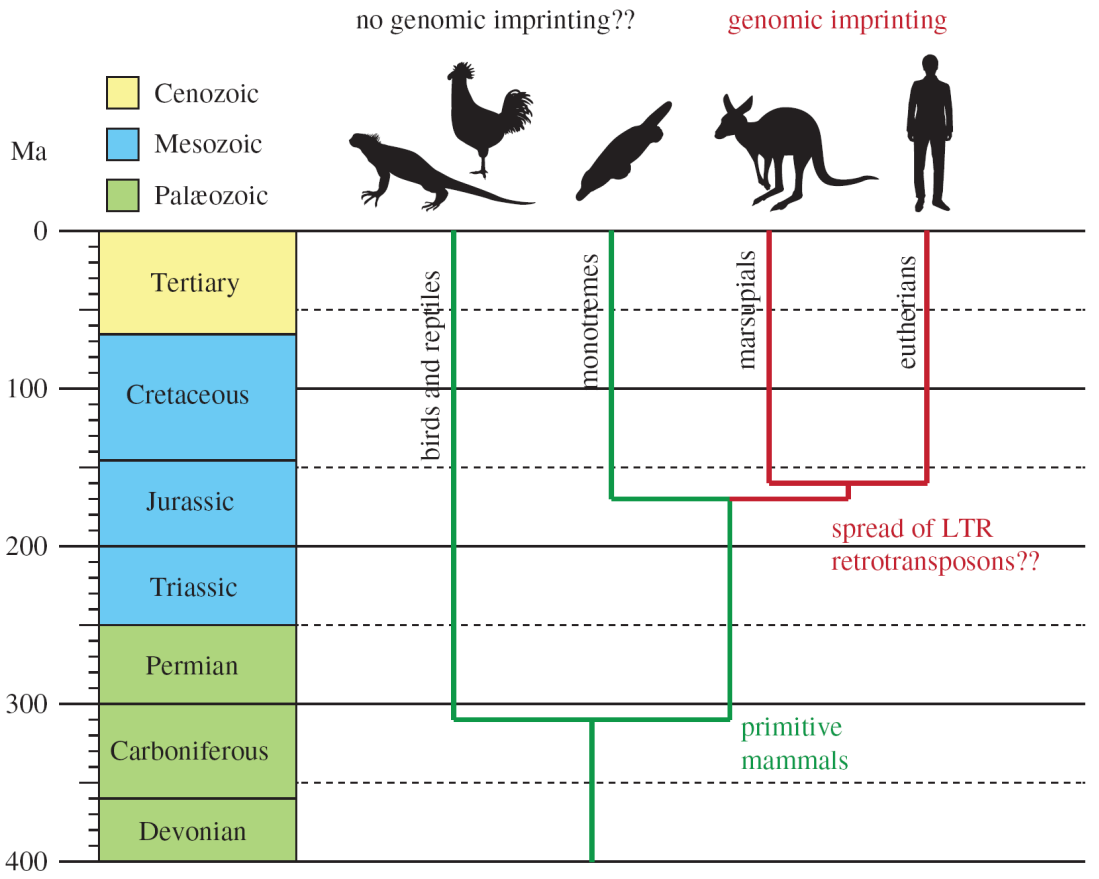


Figure 1.1 Imprinting evolution in animals appears to be restricted to mammalian lineages with a placental habit. Green lines in the cladogram denote lineages thought to lack genomic imprinting while red lines are lineages that display genomic imprinting. LTR = long terminal repeat; Ma = million years ago. Figure obtained from (Renfree et al., 2013).

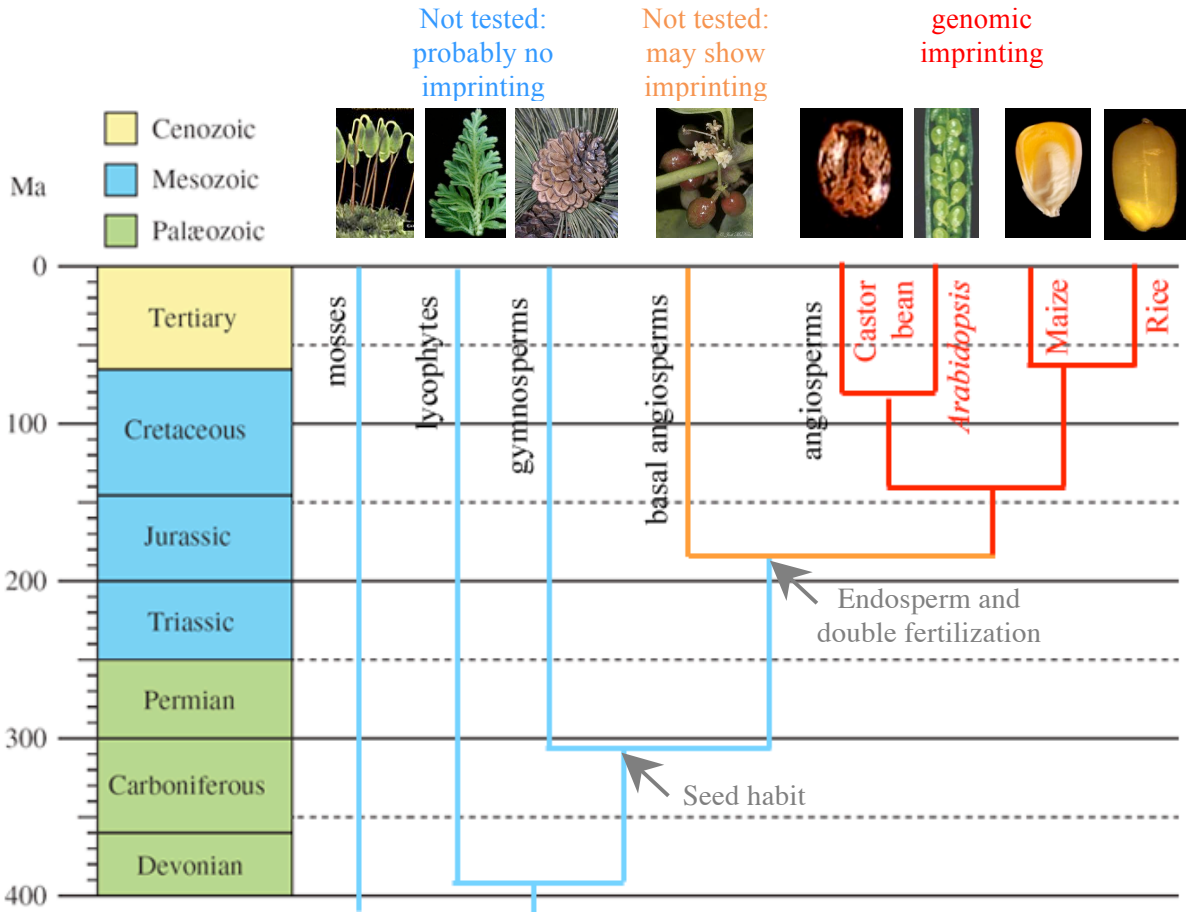


Figure 1.2 Imprinting in plant lineages is thought to be restricted to angiosperm lineages with endosperm, though other lineages are poorly sampled. Blue lines on the cladogram denote untested lineages thought to lack genomic imprinting, red lines are lineages that display genomic imprinting, and the orange line denotes currently untested lineages that are expected to show imprinting. Species surveyed for imprinting at a genome-wide scale are named on the tree (Xu et al., 2014; Hsieh et al., 2009; Waters et al., 2011; Luo et al., 2011). Arrows indicate two major events in the evolution of angiosperms. Ma = million years ago. Figure made by modifying that from (Renfree et al., 2013).

Sexual reproduction in plants and animals

Anisogamous sexual reproduction, defined by the fusion of two dissimilarly sized haploid gametes to produce a diploid offspring with potentially new genetic combinations, evolved independently in plant and animal lineages (Lehtonen & Parker, 2014). Males are, by definition, the sex that produce the smaller gametes, while females are the sex with larger gametes. The fusion of a male gamete and a female gamete, a process termed fertilization, results in the production of a single cell called the zygote. The zygote goes on to divide mitotically and form the embryo, which is the offspring at its earliest stage of development. The embryo eventually matures into a new fully-developed individual, which, in turn, goes on to produce gametes that participate in the next round of sexual reproduction.

In animals, gamete production occurs as cells in the diploid individual divide by meiosis to form single-celled haploid gametes that can be fertilized. The life cycle of plants however involves an alternation of generations between multicellular haploid and diploid states (Berger & Twell, 2011). The haploid or 'gametophyte' phase has an end goal of producing gametes (i.e. sperm and egg cells), which fuse during sexual reproduction to give rise to a diploid sporophyte. The diploid or 'sporophyte' phase has an end goal of producing haploid spores by meiosis, which in turn divide mitotically to form a gametophyte. Algae, non-vascular land plants, and the algal ancestors of modern vascular land plants all display a dominant gametophytic phase associated with vegetative growth and photosynthesis. Vascular land plants have evolved a dominant sporophytic phase for vegetative growth and photosynthesis, so the gametophytic phase is only observed in reproductive organs (Niklas, 1997). It is thought that the early success of vascular plants on dry land is attributable to advantages resulting from a dominant diploid phase of vegetative growth (Niklas, 1997).

Given that anisogamous reproduction evolved independently in mammals and flowering plants, divergent patterns of gametogenesis and zygotic development will bear differently on mechanisms, functions, and evolutionary drivers of imprinting in the two clades (Pires & Grossniklaus, 2014). With this in mind, brief summaries of gametogenesis and zygotic development are contrasted between mammals and flowering plants.

Gametogenesis in mammals

In mammals, developmental plans for the next round of sexual reproduction are laid down within less than a week after a zygote forms. At 6 days of fertilization in mice, precursor primordial germ cells in the embryo form the lineage dedicated to gamete production (MacDonald & Mann, 2014). Primordial germ cells have the potential to commit either to spermatogenesis, which is the process of producing sperm or male gametes, or to oogenesis, which is the process of producing oocytes or female gametes (Adams & McLaren, 2002). Specification of male or female fates relies on whether a 'masculinizing' environment is offered by surrounding gonadal cells in the embryo; if such an environment is provided, primordial germ cells proceed to spermatogenesis, otherwise oogenesis is the default state (Adams & McLaren, 2002).

In male mice, primordial germ cells mostly remain mitotically inactive in the embryo, only re-entering mitosis at 5 days after the birth of the individual, when prospermatogonia are formed (Ewen & Koopman, 2010). Prospermatogonia reach their potential for gamete production at puberty, when they form spermatogonia. Spermatogonia produce spermatocytes that enter into meiosis, eventually resulting in the production of haploid sperm cells (Ewen & Koopman, 2010). Female primordial germ cells, on the other hand, enter into meiosis in the 13.5 day-old mouse embryo (Ewen & Koopman, 2010). By 15.5 days, cells reach prophase I of meiosis I and then enter an arrested state. After birth, these cells, referred to as primary oocytes, remain arrested at diplotene I of meiosis I till puberty. At puberty, meiosis I is completed and a secondary oocyte is formed. The secondary oocyte proceeds through meiosis II and then arrests at metaphase II. Progression from metaphase II to anaphase II only occurs at the time of fertilization (Ewen & Koopman, 2010). Once the secondary oocyte is fertilized by a sperm cell, meiosis is completed, giving rise to a haploid ovum or egg cell that will fuse with the sperm.

Fertilization and zygotic development in placental mammals

Fertilization in mammals occurs internally, within the maternal parent's reproductive tract. In placental mammals, the resultant diploid zygote divides mitotically, and begins to differentiate into an outer trophoblast layer and an inner cell mass around day 4 after fertilization. The inner cell mass goes on to differentiate and form four tissues: the embryo proper, the amnion, the yolk sac, and the allantois. The trophoblast on the other hand goes on to form the outer layer of the placenta, an organ composed of embryonic and maternal tissues (Johnson & Selwood, 1996).

The placenta attaches the developing embryo to the wall of its mother's uterus, bringing maternal and embryonic blood supply into close contact. It grows throughout embryonic development till birth, performing key roles in facilitating the exchange of nutrients, hormones and antibodies between the maternal parent and offspring, and eliminating waste produced by the developing offspring (Frost & Moore, 2010). More imprinted genes are expressed in the placenta than in any other tissue, with perturbations of imprinting at placenta-expressed genes resulting in a range of phenotypes (Frost & Moore, 2010). A significant amount of imprinted expression also occurs in the brain, predominantly during development but also after birth (Bartolomei & Ferguson-Smith, 2011). Perturbations in imprinted expression in the embryo also result in a range of developmental disorders (Bartolomei & Ferguson-Smith, 2011).

Gametogenesis in flowering plants

Unlike in animals, plant cells are committed to reproductive fates late in development, usually in response to specific environmental and internal cues (Baurle & Dean, 2006). In angiosperm land plants, the production of male and female gametes by their respective gametophytes occurs within a reproductive organ called the flower. Floral development is initiated as the shoot apical meristem, a pool of undifferentiated dividing cells present at the apex of the plant shoot, receives cues that turn it into an inflorescence meristem (Alvarez-Buylla et al., 2010). Floral meristems that bud off of the inflorescence meristem give rise to four floral organs in concentric whorls: two outer vegetative whorls consisting of sepals and petals, and two inner reproductive whorls consisting of male stamens and one or more female carpels. Whorls are initiated beginning with outer whorls, so the carpel is the last to differentiate from the meristem.

The female gametophyte or megagametophyte develops in structures of the carpel called ovules. In-depth discussions of the structure, function, and developmental program of the female gametophyte (Drews & Koltunow, 2011; Yadegari & Drews, 2004), are summarized as follows. The process of female gamete production can be broken down into two stages: megasporogenesis and megagametogenesis. Precise developmental patterns during these two stages appear to vary among plants, but most plants show a 'Polygonum-type' pattern. This is the pattern that will be discussed as it is found in many key model plants, including maize, rice, *Arabidopsis*, soybean, tomato, and other plants that belong to their families.

Megasporogenesis is initiated in the developing ovule after the plant version of a primordial germ cell, called an archesporial cell, differentiates into a megaspore mother cell. During megasporogenesis, the megaspore mother cell undergoes meiosis to produce four haploid megaspores. At this stage, a similarity between mammalian and plant female gametogenesis is

exhibited as three of the four haploid megaspores degenerate and only one is selected to proceed on to megagametogenesis. During megagametogenesis, the megaspore undergoes three rounds of meioses to form a multicellular structure called the female gametophyte or megagametophyte. Cell walls are not formed till the last mitotic round, allowing two nuclei, one from each pole of the gametophyte, to migrate to the center and be enclosed within a single cell. The end result is a structure of seven cells: three haploid antipodal cells at the end furthest away from the entrance of the ovule, one large homodiploid central cell, and a haploid egg cell flanked by two haploid synergid cells closest to the ovule entrance. Although the egg cell is the female gamete that will genetically contribute to the next generation, the central cell also participates in a fertilization event with a second male sperm cell and makes critical contributions to the next generation, as will be discussed later.

The male gametophyte, also known as the microgametophyte or pollen, develops in the anther of a stamen (Bleckmann, Alter, & Dresselhaus, 2014; Wilson & Zhang, 2009). In brief, the process is initiated when some archesporial cells of the developing anther differentiate into pollen mother cells. Microsporogenesis involves the meiotic division of pollen mother cells into haploid microspores. Once again, as is seen in mammalian gametogenesis, all four haploid cells survive and go on to produce male gametes. Microgametogenesis proceeds as each haploid microspore undergoes two mitotic events, forming a tricellulate pollen grain with one haploid vegetative cell and two haploid sperm cells. The larger vegetative cell engulfs the two functionally-equivalent sperm cells and functions in delivering them to the female gametophyte.

Fertilization and zygotic development in flowering plants

Fertilization in plants, similar to that in animals, occurs in the female reproductive organ. When dispersed pollen lands on a compatible female reproductive organ, the vegetative cell produces a tube that guides sperm cells directly to the ovule (Bleckmann et al., 2014; Yadegari & Drews, 2004). Upon reaching the entrance to the ovule, cellular interactions cause the pollen tube to rupture and release both haploid sperm cells. 'Double fertilization' occurs when one haploid sperm cell fuses with the haploid egg cell to produce a diploid zygote, while the other haploid sperm cell fuses with the diploid central cell to produce a triploid cell that will give rise to a tissue called endosperm.

Double fertilization heralds the beginning of seed development (Bleckmann et al., 2014; Sabelli & Larkins, 2009). During seed development, the diploid zygote proliferates mitotically to form an embryo, the triploid endosperm proliferates mitotically to support and nourish the growing embryo, and the outer layers of the ovule form a protective outer covering for the seed called the seed coat. At maturation, the seed desiccates and enters a dormant state until suitable germination conditions are encountered (J. Li & Berger, 2012; Sabelli & Larkins, 2009).

It should be highlighted here that though the endosperm does not form part of the offspring plant, it is essential for proper embryo development and is analogous to the mammalian placenta in its roles (Bleckmann et al., 2014; Lafon-Placette & Kohler, 2014). The endosperm shares another similarity with the placenta by being the predominant site of genomic imprinting (Hsieh et al., 2011; Luo et al., 2011; Pignatta et al., 2014; Waters et al., 2011; M. Zhang et al., 2011). No imprinted genes have been detected in plant seedlings or adult vegetative tissues (Chodavarapu et al., 2012; Gehring, 2013; G. He et al., 2010; X. Zhang & Borevitz, 2009), and

imprinting in the embryo appears to be nonexistent or limited and more transient in nature (Del Toro-De Leon, Garcia-Aguilar, & Gillmor, 2014; Jahnke & Scholten, 2009; Luo et al., 2011; Pignatta et al., 2014).

It is also worth noting here that although endosperm development across plants shows a great deal of similarity in early stages, the role of endosperm at the time of maturation varies (J. Li & Berger, 2012). In plants such as *Arabidopsis thaliana*, much of the endosperm is consumed by late embryogenesis, and food reserves in the dormant seed are stored in the cotyledons of the embryo (J. Li & Berger, 2012). This is also the case in several other dicotyledonous plants (dicots), such as peas, beans, lentils, soybeans, peanuts and other members of the legume family, whose seeds form important food sources for humans and animals. However, in many monocotyledonous plants (monocots), including the grass species rice and maize, the endosperm plays an important role beyond early embryo development by serving as the primary storage house for nutrients as the seed matures into a quiescent state (Sabelli & Larkins, 2009). Monocot seeds also make major contributions towards human and animal nutrition, with grasses in particular producing cereal grains that feed much of the world. These and other differences in late roles for the endosperm may exert different selective pressures on individual genes across species, possibly portending differences in imprinting of homologous genes in cereals compared to *Arabidopsis* and similar plants.

Mechanisms of genomic imprinting in mammals

Models of imprinting regulation were first proposed as a result of extensive genetic studies of mouse imprinted genes in the 1990s and 2000s (Adalsteinsson & Ferguson-Smith, 2014; Ferguson-Smith, 2011). Many mammalian imprinted genes form clusters at specific loci on chromosomes, such that the imprinting of all genes in the cluster is controlled by a single distinct region with parent-of-origin-specific epigenetic marks (Ferguson-Smith, 2011). These regions, termed 'imprinting control regions', not only become associated with different epigenetic marks in the lineages leading to male and female gametes but also retain parent-of-origin specific chromatin states after fertilization by escaping chromatin reshaping mechanisms in the embryo (Proudhon et al., 2012). Epigenetic marks set up before the formation of the zygote are referred to as primary imprints. In theory, a primary imprint may be set up either during gametogenesis or sometime during the fertilization process but before fusion of the nuclei of male and female gametes (Ferguson-Smith, 2011). Parent-of-origin-dependent differences that appear in the zygote or later stages, once maternally- and paternally-inherited chromosomes are in the same nucleus, are referred to as secondary or somatic imprints. Such secondary imprints rely on the presence of a primary imprint but add new dimensions to parent-of-origin-specific gene regulation.

DNA methylation

DNA methylation was the first epigenetic mark to be associated with mammalian imprinting (Bartolomei, Webber, Brunkow, & Tilghman, 1993; Ferguson-Smith, Sasaki, Cattanach, & Surani, 1993; Stoger et al., 1993) and is still the only primary imprint that has been discovered in mammals (MacDonald & Mann, 2014). In order for sex-specific methylation marks to be set up during gametogenesis, an erasure of maternal and paternal imprints inherited from the previous generation occurs specifically in precursor primordial germ cells, with initial

processes starting at just 7.5 days after fertilization in mice (Saitou, Kagiwada, & Kurimoto, 2012). Active and passive DNA demethylation mechanisms deplete primordial germ cells of methylation till, at 13.5 days after fertilization in mice, the majority of imprinted regions are demethylated (Guibert, Forne, & Weber, 2012; Hackett et al., 2013; Kagiwada, Kurimoto, Hirota, Yamaji, & Saitou, 2013).

Although the meiotic divisions associated with spermatogenesis do not take place till puberty, male germ cell-specific methylation events begin at 16.5 days after fertilization, long before even prospermatogonia form (Kobayashi et al., 2013). The methylation of all three known paternally-methylated imprinting control regions is completed by 18.5 days after fertilization (Kato et al., 2007). By contrast, although the meiotic divisions of oogenesis are initiated in the developing embryo, female primordial germ cells remain hypomethylated through meiotic arrest, with recent studies suggesting that some demethylation continues past 16.5 days after fertilization (Kobayashi et al., 2013).

DNA methylation events specific to the female germline begin to take place only after puberty, continuing up until the release of the secondary oocyte from the ovary (Lucifero, Mann, Bartolomei, & Trasler, 2004). In either case, by the time fertilization occurs, the germline has gone through significant epigenetic modification that is specific to each sex, though it is interesting to note that most imprinting control regions are maternally methylated and only a handful are paternally methylated (Bartolomei & Ferguson-Smith, 2011).

In order to function as a primary imprint, a DNA methylation mark established before fertilization must be one of the minority that survive the wide-scale epigenetic reprogramming that begins just after fertilization (Mayer, Niveleau, Walter, Fundele, & Haaf, 2000; Oswald et al., 2000) and continues until the embryo implants into the wall of its mother's uterus. During this process, the imprint is dynamic: enlarging, contracting, and shifting position until its final state in the embryo (Tomizawa et al., 2011). Imprints that regulate unclustered imprinted genes tend to occur in the promoter region (Bartolomei & Ferguson-Smith, 2011).

Interestingly, the promoter region is also the location of most maternal-specific methylation, while paternal-specific methylation tends to occur in intergenic regions (Bartolomei & Ferguson-Smith, 2011; Ferguson-Smith, 2011). Intergenic and promoter imprinting control regions appear to regulate imprinting through different mechanisms (Bartolomei & Ferguson-Smith, 2011; Bourc'his & Bestor, 2006), so it is possible that these sex-specific differences in the mechanism of action of the imprint may reflect different mechanisms for establishing the primary imprint or different roles for the imprints (Bourc'his & Bestor, 2006).

Another interesting finding is that some critical genes imprinted during embryogenesis lose imprinting later on in development and that this is associated with changes in DNA methylation at imprinting control regions (Ferron et al., 2011). At the *Dlk1* locus, the effective epigenetic resetting of this gene is required for higher expression in order to fulfill its postnatal role. This once again highlights the importance of DNA methylation as an indicator and regulator of the expression state of imprinted and non-imprinted genes.

Histone modifications

Selective histone remodeling of paternally-inherited DNA immediately after fertilization results in another layer of parent-of-origin-specific epigenetic differences in mammals. DNA in the sperm cell, unlike DNA in other cells, is not predominantly packaged with histones. Instead, protamines, small arginine-rich proteins synthesized late in spermatogenesis, are used to tightly condense most (90 to 99% depending on the species) of the genome into a silenced state (Balhorn, 2007). Remodeling processes in the zygote swap these protamines for H4 acetylated histones (Adenot, Mercier, Renard, & Thompson, 1997) and H3 histones unmarked by lysine 27 trimethylation (Erhardt et al., 2003) and lysine 9 di- and trimethylation (Liu, Kim, & Aoki, 2004; Santos, Peters, Otte, Reik, & Dean, 2005). Maternally-inherited DNA retains H3 histone modifications, resulting in differences between maternal and paternal genomes that last till the four cell stage.

After the 4-cell stage, parent-of-origin-specific histone modifications mark sites of differential DNA methylation, probably guided by DNA methylation differences themselves. Alleles with DNA methylation are enriched for histone 3 lysine 9 di- and trimethylation (Messerschmidt et al., 2012) and histone 4 lysine 20 trimethylation (McEwen & Ferguson-Smith, 2010) while alleles lacking DNA methylation are enriched for histone 3 lysine 4 trimethylation (McEwen & Ferguson-Smith, 2010). These differential histone marks no doubt contribute to the silencing of coding and non-coding transcripts at methylated imprinting control regions and the activation or continued expression of transcripts at unmethylated imprinting control regions. Other active and repressive histone marks are noted to correlate with expressed and unexpressed alleles, but there is less evidence that these help form the imprint and more evidence that they are by-products of developmental expression states (Ferguson-Smith, 2011)

Non-coding RNAs: long non-coding RNAs and small RNA

Imprint establishment at one paternally-imprinted mouse locus, *Rasgrf1*, relies on the Piwi-interacting RNA pathway, which targets transcriptional and post-transcriptional gene silencing through small (26 – 31 nt) non-coding RNAs (Watanabe et al., 2011). This pathway is not required for imprint establishment at other paternally-imprinted loci (Watanabe et al., 2011), leading to the important conclusion that different epigenetic pathways may contribute to imprint establishment at different sites (MacDonald & Mann, 2014).

Two of the other three known paternally-imprinted loci in mice show a correlation between transcript production and the initiation of DNA methylation acquisition, suggesting that transcription might target repressive chromatin modifiers, including DNA methyltransferases, to imprinted regions (Henckel, Chebli, Kota, Arnaud, & Feil, 2012). A similar phenomenon is seen on the maternal side at several genic loci in developing oocytes (Chotalia et al., 2009). However, it should be noted that although oocytes do not transcribe through paternally-imprinted regions, male primary germ cells transcribe through maternally-imprinted regions and yet maternally-imprinted regions remain unmethylated on the paternal allele. This may be due to other epigenetic marks, such as histone 3 lysine 4 trimethylation, that might be present in male primary germ cells at some loci but not at paternally-imprinted ones (Henckel et al., 2012).

Non-coding RNAs are not just involved in establishing imprints but also in mediating the silencing effects of imprints post-fertilization, particularly at imprinting clusters. Maternal expression of the *Igf2r* gene is associated with intronic methylation that suppresses the expression of a long noncoding RNA, *Airn* (Lyle et al., 2000). The paternal allele lacks intronic methylation, allowing expression of *Airn* and leading to the repression of *Igf2r* and two other downstream genes. Current evidence suggests that *Airn* acts by recruiting a histone 3 lysine 9 methyltransferase to silence genes in cis (Nagano et al., 2008). Cis-acting regulatory non-coding RNAs have also been identified at three other imprinted clusters, where once again maternal methylation is associated with repression of the non-coding RNA (Chotalia et al., 2009; Fitzpatrick, Soloway, & Higgins, 2002; Horsthemke & Wagstaff, 2008).

CTCF-dependent insulator regions

CTCF is a zinc-finger insulator protein which preferably binds to unmethylated DNA (Ong & Corces, 2014) and has been implicated in imprinting control at some but not all intergenic imprinting control regions (Bartolomei & Ferguson-Smith, 2011). *H19* and *Igf2* imprinted genes occur within 90 kb of each other and share enhancer elements located downstream of *H19*. The imprinting control region upstream of the *H19* gene contains several CTCF binding sites, and paternal-specific DNA methylation results in CTCF preferentially binding to the maternal allele (Bell & Felsenfeld, 2000; Hark et al., 2000). On the maternal allele, CTCF presence blocks enhancers from stimulating *Igf2* expression, so expression instead occurs from the *H19* allele. On the paternal allele, CTCF absence allows enhancers to stimulate *Igf2* expression, resulting in maternal expression of *H19* and paternal expression of *Igf2*. The critical involvement of CTCF at some imprinting control regions but not others once again highlights how imprinting is mediated through different epigenetic mechanisms at different loci, and that the study of imprinting has potential to yield great general insight into mechanisms of epigenetic regulation (Adalsteinsson & Ferguson-Smith, 2014; Ferguson-Smith, 2011).

Mechanisms of genomic imprinting in plants

The convergent evolution of imprinting in placental mammals and flowering plants relies upon epigenetic mechanisms that show certain degrees of conservation and divergence across eukaryotes (X. J. He, Chen, & Zhu, 2011; Keller & Buhler, 2013; Ong & Corces, 2009; Pinney, 2014; Saze, Tsugane, Kanno, & Nishimura, 2012; Zemach, McDaniel, Silva, & Zilberman, 2010). Among notable similarities are the conserved role of the DNMT family in mediating DNA methylation, the conserved role of several histone modifying complexes such as the polycomb group comples, and the presence of small RNA pathways where conserved dicer and argonaute proteins mediate gene silencing at transcriptional and post transcriptional levels. Differences that distinguish plants from animals include the absence of CTCF-dependent insulators (Ong & Corces, 2009), the presence of an RNA-directed DNA methylation pathway that involves plant-specific RNA polymerases (Law & Jacobsen, 2010), and the prevalence of cytosine DNA methylation in non-CG contexts (Pinney, 2014).

Differences in mechanisms of imprinting in plants and animals will thus not only reflect differences in developmental programs in each clade, but also be influenced by divergent epigenetic regulation. With the dearth of data from technically-challenging investigations of parent-of-origin-specific epigenome dynamics at various time points during plant gametogenesis,

fertilization, and zygotic development, it is unclear whether DNA methylation is as dynamic during plant reproduction as it is in animals. Reports of pathways that reinforce transposon silencing in the plant germline (Calarco et al., 2012; Ibarra et al., 2012) suggest that although some reprogramming does occur during plant reproduction, it occurs on a much smaller scale compared to that in animals. Much is left to be discovered about how imprinted expression is mediated on plant chromosomes.

DNA methylation

The first regulator of imprinting identified in plants was the DEMETER (DME) DNA glycosylase, which acts in the central cell but not sperm cells, resulting in maternal-specific DNA demethylation in the endosperm (Choi et al., 2002; Ibarra et al., 2012; T. Kinoshita et al., 2004). DNA demethylation by DME has been associated with both maternal-specific gene activation (Hsieh et al., 2011; T. Kinoshita et al., 2004) and maternal-specific gene repression (Hsieh et al., 2011; Villar, Erilova, Makarevich, Trosch, & Kohler, 2009). It is currently unknown how DME is targeting to specific sites, though studies reveal that DME demethylated regions are enriched for small euchromatic transposable elements (Ibarra et al., 2012).

The finding that most parent-of-origin-specific differences across the endosperm genome disappear in the *dme* mutant (Ibarra et al., 2012) suggests that, unlike in mammals, either only a small fraction of inherited DNA methylation is changed after male and female germlines differentiate from the floral meristem, or that DNA methylation alterations after differentiation from the floral meristem proceed simultaneously in male and female germlines. Mutations in the MET1 DNA methyltransferase result in a loss of imprinting (Hsieh et al., 2011; Xiao et al., 2003), suggesting that it plays a role in maintaining methylation in the germline, and is responsible for the paternal methylation of loci that are maternally demethylated by DME.

It is interesting to note that the activity of DEMETER is confined to the central cell of the female gametophyte (Choi et al., 2002; Ibarra et al., 2012; T. Kinoshita et al., 2004), even though one of the two central cell nuclei is separated from the egg cell by only one mitotic division (Yadegari & Drews, 2004). It may be that there is a greater need for mammalian germ cell lineages to undergo more intense epigenetic reprogramming, in order to erase maternal and paternal imprints inherited from the previous generation. The lack of imprinting in embryos, plant seedlings, and adult vegetative tissues (Chodavarapu et al., 2012; Gehring, 2013; G. He et al., 2010; X. Zhang & Borevitz, 2009) suggests that the meristematic tissue that gives rise to floral reproductive whorls is devoid of any imprints.

Another explanation for a more extensive reprogramming of DNA methylation in mammals than in plants could be that differences in germline developmental programs between plants and mammals allow more time for a dramatic reshaping of the genome in mammals compared to plants. Plant male and female organs are found within the same flower and differentiate from the same pool of meristematic cells within a few weeks (Smyth, Bowman, & Meyerowitz, 1990).

Histone modifications

The involvement of the polycomb group in mediating imprinting was implicated by the finding that MEDEA, a member of the polycomb group complex, is required for the paternal expression of the *PHERES1* gene in *Arabidopsis* (Kohler, Page, Gagliardini, & Grossniklaus, 2005). Since then, the polycomb complex has been implicated in the imprinted regulation of other *Arabidopsis* genes (Hsieh et al., 2011), including maternal expression of *MEDEA* itself (Gehring et al., 2006; Jullien, Katz, Oliva, Ohad, & Berger, 2006). It is interesting to note that almost all paternally-expressed imprinted genes appear to be regulated by the polycomb group (Hsieh et al., 2011), and that polycomb activity on the materally-inherited is more often than not also associated with differential DNA methylation marks (Gehring et al., 2006; Hsieh et al., 2011; Jullien et al., 2006; Kohler et al., 2005). Several studies have now shown that polycomb activity, detected through the presence of histone 3 lysine 27 trimethylation marks, is anticorrelated with DNA methylation (Deleris et al., 2012; M. Du, Luo, Zhang, Finnegan, & Koltunow, 2014; Weinhofer, Hehenberger, Roszak, Hennig, & Kohler, 2010). It is unclear whether the histone 3 lysine 27 trimethylation mark is a primary imprint at some loci, and whether other histone modifications may also form primary or secondary imprints.

Long non-coding RNAs

At the moment, no specific long non-coding RNAs are known to be required for imprinted gene expression in the endosperm, but it is likely that such RNAs exist. The best studied target of polycomb-mediated histone modification in *Arabidopsis*, the *FLC* locus, is associated with a long non-coding RNA (Csorba, Questa, Sun, & Dean, 2014). In addition, several non-coding RNAs in mammals have been linked to targeting of the polycomb group in cis and in trans (Brockdorff, 2013), suggesting it is a widespread mechanism of polycomb targeting. It is possible that imprinted non-coding RNAs and alternatively-spliced regions observed in rice and maize (Luo et al., 2011; M. Zhang et al., 2011) may later be identified to be associated with the targeting of repressive chromatin modifiers in a manner that mediates imprinted expression at coding sequences.

Small RNAs

It has been shown that the activity of small RNAs in the germline is required for proper methylation at the paternal allele of several imprinted loci (Vu et al., 2013), suggesting that small RNAs play a similar role to MET1 in maintaining methylation in the germline. In plants, 24-nt small RNAs are associated with targeting of the RNA-dependent DNA methylation pathway to loci with particular epigenetic characteristics, in order to mediate transcriptional silencing (Law & Jacobsen, 2010). The role of small RNAs in forming an imprint or mediating silencing at imprinted regions is currently poorly understood. Reports of the presence of large populations of maternally-biased but not paternally-biased 24-nt small RNAs in the young *Arabidopsis* seed (Mosher et al., 2009) have lead to investigations of the role of small RNAs in mediating imprinted gene expression (Pignatta et al., 2014), and the role of imprinted small RNAs in mediating seed development (Lu, Zhang, Baulcombe, & Chen, 2012; Mosher et al., 2009).

Initial work suggested that the massive accumulation of these maternally derived small RNAs in *Arabidopsis* is unaffected by mutations in any of the DNA- and histone-modifying

enzymes that are known to regulate imprinted gene expression, including DME (Mosher et al., 2011), leaving unresolved the mechanism by which they are generated. However, recent work seems to indicate that many of these maternally-biased small RNAs might originate in the seed coat, which lies in close proximity to the *Arabidopsis* endosperm at this stage and forms a wide area of contact with it (Pignatta et al., 2014). The same study also suggests that truly imprinted *Arabidopsis* small RNAs show both maternal and paternal bias, with paternally-biased small RNAs, which are highly unlikely to be due to contamination of the endosperm with paternally-derived tissue, being enriched at the promoters of maternally expressed imprinted genes. A similar correlation of imprinted small RNA with the silenced allele of imprinted genes was also observed in maize (Xin et al., 2014).

The distinct retention of parent-of-origin-specific differences in expression, despite the production of trans-acting small RNAs that can silence alleles, remains a mystery. It has been proposed that this might be due to an unusual arrangement of chromatin in the endosperm compared to other tissues (Pignatta et al., 2014), and there is some support for this theory (Baroux, Pecinka, Fuchs, Schubert, & Grossniklaus, 2007). It is also possible that imprinted small RNAs are only seen in the subset of cells in the endosperm where a resetting of the imprint is being initiated, and that end result of such resetting would be biallelic expression or biallelic repression. Evidence in mammals shows that proper development may require the monoallelic expression of a gene early during development and biallelic expression later on (Ferron et al., 2011), and it is possible that similar dynamics might be at work in plants. However, as the endosperm is a terminal tissue that undergoes programmed cell death at the end of seed development (J. Li & Berger, 2012; Sabelli & Larkins, 2009), it is less likely that it should be under selective pressure to revert back to biallelic expression.

Theories of evolutionary forces driving imprinting in mammals and plants

Theories of imprinting evolution have been widely contested, debated, and misunderstood for several years now (D. Haig, 2014; Patten et al., 2014; Pires & Grossniklaus, 2014). All agree that, at first glance, the monoallelic expression of a gene seems to be disadvantageous as deleterious mutations cannot be complemented by a functional homologous locus. For example, mutations in the *Arabidopsis* maternally expressed *MEDEA* gene result in aborted seeds despite seeds inheriting a functional copy of the gene from the paternal parent (Grossniklaus, Vielle-Calzada, Hoeppner, & Gagliano, 1998). However, the widespread prevalence of genomic imprinting in both placental mammals and angiosperm land plants, despite the associated fitness cost, suggests that imprinting must increase fitness in other respects. Various hypotheses have been proposed and here I will discuss some of the most popular and well-supported ones that are relevant to plants.

Kinship theory or parental conflict hypothesis

David Haig and Mark Westoby developed the "parental conflict hypothesis" of imprinting while pondering the relevance of imprinted gene expression in triploid endosperm (David Haig & Westoby, 1989). The hypothesis predicts that, in a situation where the female parent contributes more resources to offspring development than the male parent and can bear the offspring of multiple males, the overall fitness of paternally-inherited DNA is increased if the offspring thrives at the cost of other offspring borne by the female parent, while the fitness of

maternally-inherited DNA is increased if the female is able to have as many successful offspring as possible. These evolutionary pressures result in paternal expression of genes that promote nutrient acquisition in the offspring and maternal expression of genes that restrict nutrient allocation to any one offspring and instead moderate nutrient allocation to all current and future offspring.

This is currently the most popular theory of imprinting in both mammals (Ferguson-Smith, 2011; Frost & Moore, 2010) and plants (Huh et al., 2008; Kohler & Weinhofer-Molisch, 2009; Pires & Grossniklaus, 2014), for several reasons. Firstly, the evolution of imprinting in each case appears to be linked with the evolution of the 'placental habit', i.e. the evolution of the placenta in mammals (Renfree et al., 2013) and the evolution of the endosperm in plants (Pires & Grossniklaus, 2014). As both the placenta and endosperm result in a more intimate dependency of the developing embryo on the maternal parent, parental conflict is intensified compared to lineages with a less intimate interaction (Crespi & Semeniuk, 2004).

Other support for parental conflict as a driving force in the evolution of mammalian imprinting comes from findings that implicate the majority of imprinted genes in functions related to nutrient acquisition from the maternal parent, including nutrient transfer from the placenta, embryo growth, brain development, and postnatal energy homeostasis (Bartolomei & Ferguson-Smith, 2011). Prime examples of imprinted genes that fit this hypothesis are paternally expressed *Igf2*, which promotes growth, and maternally expressed *Igf2r*, which appears to moderate growth based on knock out studies (D. Haig & Graham, 1991). In plants, interploidy crosses in *Arabidopsis thaliana* that increase paternal gene dosage result in larger seeds while those that increase maternal gene dosage result in smaller seeds (Scott, Spielman, Bailey, & Dickinson, 1998). Later on, specific imprinted genes were also identified to play roles in endosperm growth, including the examples of the maternally expressed members of the polycomb group complex *(Arabidopsis MEDEA* and *FIS2*, and maize *fie1* and *fie2*), that appear to restrict growth (Grossniklaus et al., 1998; Hermon, Srilunchang, Zou, Dresselhaus, & Danilevskaya, 2007; Kiyosue et al., 1999; Luo, Bilodeau, Dennis, Peacock, & Chaudhury, 2000) and the paternally expressed *Arabidopsis PHERES1* gene that appears to promote growth (Makarevich, Villar, Erilova, & Kohler, 2008; Pires & Grossniklaus, 2014).

However, the parental conflict theory does not explain all observations of imprinting in plants. For example, if nutrient acquisition and growth of the developing offspring were the main drivers of imprinted expression, one might expect to observe imprinting in the developing embryo in addition to extraembryonic nutritive tissues, as is seen in mammals. By contrast, imprinting in the embryo is nonexistent, or limited and transient (Del Toro-De Leon et al., 2014; Jahnke & Scholten, 2009; Luo et al., 2011; Pignatta et al., 2014). It is possible that this difference between plants and mammals may reflect underlying developmental differences in the extent to which different extraembryonic tissues regulate embryo growth.

Another argument against the parental conflict model is the presence of imprinted expression in self-fertilizing plants such as *Arabidopsis thaliana* and domesticated rice cultivars. As maternal and paternal genomes are related in these cases, it would be expected that the genomes would work synergistically rather than antagonistically. However, it is possible that the presence of genomic imprinting may be the signature of a recent ancestral out-crossing state

(Spillane et al., 2007) or that imprinting evolved due to intense conflict in a distant ancestral population but has been exapted for other purposes in contemporary lineages (Ferguson-Smith, 2011).

Coadaptation hypothesis

A further observation inconsistent with parental conflict is the dominance of maternally-expressed genes in some species, including *Arabidopsis* (Gehring, Missirian, & Henikoff, 2011; Hsieh et al., 2011; McKeown et al., 2011), and the presence of maternally-expressed genes that do not appear to restrict nutrient allocation to the offspring (Costa et al., 2012). This has led to the proposal that imprinting of some maternally expressed genes may be driven by the need for efficient communication between the maternal parent and offspring while the offspring is reliant on the maternal parent during development (Wolf & Hager, 2006). This 'coadapatation' theory of imprinting does not appear to be widely applicable in all plant species though. Maize, for example, appears to possess a larger number of paternally expressed genes than maternally expressed genes (Waters et al., 2011; M. Zhang et al., 2014). However, it could be predicted that shifts in reproductive strategies should result in a different distribution of evolutionary pressures, allowing different evolutionary scenarios to be the most relevant for different species.

Dosage hypothesis

In mammals, many imprinted genes appear to be dosage sensitive (Bartolomei & Ferguson-Smith, 2011). This supports the parental conflict model of imprinting evolution, as most conflict plays out in the context where differences in dosage cause different phenotypic effects (Haig, 2014; Patten et al., 2014). However, it has also been proposed by some that genomic imprinting may have evolved in response to the need for a strict control over gene expression at critical developmental stages (Ferguson-Smith, 2011; Dilkes & Comai, 2004). This is supported by findings that many imprinted genes are often expressed in an imprinted manner in precise tissues at precise times during development (Ferguson-Smith, 2011). An example is provided by the paternally expressed gene *Dlk1*, which plays roles in development in the embryo and also in neural stem cell function (Ferron et al., 2011). Development of the embryo is perturbed when the *Dlk1* gene is expressed from two copies instead of just one paternal copy (Takahashi, Kobayashi, & Kono, 2010), while proper neurogenesis after birth requires biallelic expression and will be perturbed if only one allele is expressed (Ferron et al., 2011). A dosage hypothesis of imprinting evolution has advantages over the parental conflict model when trying to explain moderate or incomplete parental bias, as parental conflict is predicted to drive imprinting to the complete silencing of one allele (Wilkins & Haig, 2003). This advantage of the dosage hypothesis might be particularly relevant to plant endosperm, where bias is not always completely maternal or completely paternal (Gehring, 2013), and gene expression proceeds with the challenge of regulating three alleles for every gene.

Post-hybridization barrier effects on speciation

It has also been proposed that imprinting may play a role in outcrossing interactions and hybridization, especially between different ploidies (Sekine et al., 2013). This presents another way in which it might be selected in self-fertilizing plants, as it contributes to speciation events that might reduce intraspecific competition. Post-hybridization effects fit in very closely with the

dosage hypothesis of genomic imprinting, as the genes resulting in a barrier to hybridization would have to exhibit dosage sensitivity (Birchler, 2014; Schatlowski & Kohler, 2012). It has been proposed that the early rapid speciation events associated with the rise of angiosperms as the dominant plant lineage might be related to polyploidization-related barriers (Schatlowski & Kohler, 2012), suggesting an important historic outcome of imprinted expression.

Genome defense in the female gametophyte

A last possibility worthy of discussion is the scenario where imprinted expression is not driven by strong selective pressures in plants, but is instead the by-product of critical processes occurring in the gametophyte. DME, the DNA glycosylase whose activity in the female cell lays down primary imprints for both maternally and paternally expressed genes, is also active in the vegetative nucleus of the pollen grain (Schoft et al., 2011). The vegetative cell does not fuse with any female gametes, so the role of vegetative cell-specific demethylation is clearly not the generation of primary imprints for monoallelic expression. DME-mediated demethylation in the vegetative cell has instead been linked to the generation of small RNAs that direct methylation in the sperm cell (Calarco et al., 2012; Ibarra et al., 2012), thus performing the important role of silencing transposons in the germline. It has been proposed that the central cell may play a similar role in the silencing of transposons in the female germline (Ibarra et al., 2012), though it is unclear whether small RNA-mediated gene silencing pathways are active in the early gametophyte (Jullien, Susaki, Yelagandula, Higashiyama, & Berger, 2012; Vu et al., 2013).

However, it should be noted that a role for DME-mediated demethylation in gametophytic genome defense is not incompatible with a role for it in evolutionarily adaptive imprinted expression. Evidence in both mammals and plants indicates that mechanisms behind imprinted expression probably evolved from those aimed at silencing transposable elements in the germline (Barlow, 1993; Ibarra et al., 2012; H. Jiang & Kohler, 2012). DNMT3L, a regulatory factor for the *de novo* DNA methyltransferase that establishes primary imprints in mice, is also essential for retrotransposon repression in the germline, suggesting both mechanisms may be linked (Bourc'his & Bestor, 2004; Bourc'his, Xu, Lin, Bollman, & Bestor, 2001; Jia, Jurkowska, Zhang, Jeltsch, & Cheng, 2007). It is also interesting to note that two imprinted genes required for proper placental development evolved from retrotransposons (R. Ono et al., 2006; Sekita et al., 2008), further linking the development of the placenta and imprinted expression to a burst of LTR retrotransposon activity that occurred around the same time (Renfree et al., 2013). It is tantalizing to speculate that endosperm development and plant imprinting may also be correlated with significant alterations to the genome and epigenome induced by transposable elements.

However, at this stage, it appears that the current limited evidence is unable to support any one particular hypothesis of imprinting over other plausible scenarios. This highlights the need for additional work to contribute towards a better understanding of genomic imprinting in plants. Through an elucidation of the nature of maternal and paternal genome interactions through imprinting mechanisms, and an accurate assessment of the general functions of imprinted genes, we will gain a better appreciation of the significance of genomic imprinting as a unique pattern of gene regulation observed at the most tender stage of life.

Chapter II

Imprinted expression of genes is associated with localized hypomethylation of the maternal genome in rice endosperm

Most of the following chapter has been published as part of a peer reviewed article in the Proceedings of the National Academy of Sciences, USA:

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D (2013). Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proceedings of the National Academy of Sciences, USA, 110* (19): 7934-7939. doi: 10.1073/pnas.1306164110

Contributions:

Pamela Ronald, Robert Fischer, and Daniel Zilberman conceived and designed the experiment, obtained NSF funding, and provided discussion and advice at various stages. Randy Ruan performed rice crosses and assisted in dissecting and harvesting rice endosperm and embryo. Toshiro Nishimura wrote and modified Perl Scripts for data analysis and assisted with some of the analysis. Manoj and Rita Sharma coordinated the resequencing of the Kitaake rice cultivar and identification of single nucleotide polymorphisms between Kitaake and Nipponbare rice cultivars with staff at the Joint Genome Institute. I contributed to experimental design, dissected and harvested rice endosperm and embryo, and performed all molecular biology experiments and most data analysis.

Sequencing data are deposited in GEO with accession number GSE44898.

The result pertaining to the conservation of patterns of maternal hypomethylation around imprinted genes between *Arabidopsis* and rice was published as part of a peer reviewed article in Science:

Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, Rojas D, Fischer RL, Tamaru H, Zilberman D (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science, 337* (6100): 1360-1364.

Contributions:

I analyzed *Arabidopsis thaliana* endosperm DNA methylation data generated by Christian Ibarra and Tzung-Fu Hsieh in order to identify differentially-methylated regions (DMRs), plot their

distribution within 5 kb of gene transcription start and termination sites, and determine whether imprinted genes showed preferred patterns of DMR localization.

The result pertaining to the default silencing of maternally expressed imprinted genes and default activation of paternally expressed imprinted genes in rice was made possible by data available after our previous publication was published, and so will be submitted for publication as part of our next peer-reviewed article.

Introduction

DNA methylation is a covalent modification of cytosine observed across the tree of life (Zemach & Zilberman, 2010). In plants, methylation is mediated by distinct enzymatic systems in the CG, CHG, and CHH contexts (where H is A, C, or T), and regulates gene expression and transposon repression (X. J. He et al., 2011). The establishment of plant DNA methylation in all sequence contexts and a significant portion of maintenance of CHH methylation is mediated by a specialized branch of the RNA interference pathway that generates nuclear-targeted 24-nt small RNA (sRNA) molecules (Law & Jacobsen, 2010; H. Zhang & Zhu, 2011). Plants also possess DEMETER (DME) family DNA glycosylases that actively demethylate DNA by excising 5-methylcytosine in all sequence contexts (Morales-Ruiz et al., 2006; Penterman et al., 2007), with a preference for relatively euchromatic transposable elements (Ibarra et al., 2012).

Homologs of DME are expressed in various tissues of the model dicot *Arabidopsis thaliana* throughout development (Zhu, 2009), but DME expression is most prominent in the so-called gamete companion cells: the central and vegetative cells (Choi et al., 2002; Schoft et al., 2011). The vegetative cell forms the pollen tube that delivers two sperm cells, one of which fuses with the haploid egg cell to form the embryo, while the other fuses with the adjacent diploid central cell to form the nutritive triploid endosperm. Together with a maternally-derived seed coat, the endosperm and embryo form a seed. DME activity in the central cell is essential for the specific hypomethylation seen in maternally inherited endosperm chromosomes (Ibarra et al., 2012), which establishes parent-of-origin-specific (imprinted) gene expression patterns that are crucial for seed development (Bauer & Fischer, 2011; Choi et al., 2002).

Rice, a monocot that diverged from *A. thaliana* roughly 150 million years ago (Chaw, Chang, Chen, & Li, 2004), possesses marked decreases in CG methylation at specific sites throughout the genome that correlate with endosperm-specific gene expression (Zemach, Kim, et al., 2010). However, we currently do not know whether the localized endosperm CG hypomethylation in rice is maternal-specific and whether it correlates with imprinted gene expression. Here, we show that hypomethylation in rice endosperm occurs specifically on maternally inherited chromosomes, is preferentially associated with imprinted genes, and is enriched within regions of open chromatin, suggesting that DME-family mediated demethylation is a conserved feature of flowering plant reproduction.

Results

Resequencing of the Kitaake cultivar

To examine parent-of-origin-specific DNA methylation and sRNA expression in rice endosperm, we sequenced the genome of the Kitaake cultivar of japonica rice, achieving 79-fold depth of coverage. Kitaake is a photoperiod-insensitive relative of Nipponbare with a rapid life cycle compared to other rice varieties (about 9 weeks from seed to seed) that facilitates genetic experiments. Using the reference MSU version 6.0 genome sequence of the closely related (Nasu et al., 2002) Nipponbare rice cultivar (Ouyang et al., 2007; International Rice Genome Sequencing Project, 2005), we mapped Kitaake reads, which covered 97% of the Nipponbare genome. We identified 169,819 single nucleotide polymorphisms (SNPs) between Nipponbare and Kitaake, allowing us to resolve the parental origin of sequence reads from F1 hybrids of these cultivars.

The maternal rice endosperm genome shows strong site-specific DNA hypomethylation

We performed bisulfite sequencing of endosperm and embryo harvested from F1 seeds of Nipponbare and Kitaake reciprocal crosses (Table 2.1). In both reciprocal crosses, maternal and paternal genomes in 7- to 8-day old rice embryos display similar methylation patterns in all sequence contexts (the average of reciprocal crosses is shown in Fig. 2.1A-C and 2.2A-F; results for individual crosses, which are very similar, are shown in Fig. 2.3 and 2.4). However, the maternal genome of 7- to 8-day old rice endosperm is slightly globally hypomethylated and strongly hypomethylated at specific sites in the CG context with respect to the paternal endosperm genome (Fig. 2.2A-B and 2.1D; consistent individual cross data shown in Fig. 2.3A-B and 2.4A-B), which has similar methylation patterns to both parental complements of the embryo genome (Fig. 2.2A-B and Fig. 2.5A). Thus, the CG hypomethylation of the maternal endosperm genome apparently fully accounts for the localized CG methylation differences we previously reported between rice endosperm and embryo (Fig. 2.5B) (Zemach, Kim, et al., 2010), and mirrors the site-specific, DME-mediated CG demethylation of the maternal genome in *A. thaliana* (Ibarra et al., 2012).

CHG and CHH methylation of both maternal and paternal genomes is lower in the endosperm than in the embryo (Fig. 2.2C-F, 2.3C-F, 2.4C-F and 2.5C-F), consistent with our earlier report (Zemach, Kim, et al., 2010). The maternal endosperm genome is modestly globally hypermethylated in the CHG context compared to the paternal genome (black trace in Fig. 2.1E), but loci with strong maternal CG hypomethylation are also maternally hypomethylated in non-CG contexts (red traces in Fig. 2.1E-F), as they are in *A. thaliana* (Ibarra et al., 2012). This indicates that the localized demethylation affects all sequence contexts, strongly implicating a DME-family glycosylase.

Table 2.1 Coverage and mean DNA methylation in CG, CHG, and CHH contexts for libraries that were bisulfite-sequenced. Chloroplast CHH methylation is a measure of cytosine non-conversion and other errors. M/P = maternal/paternal; the expected ratio is 1 for embryo and 2 for endosperm.

Library	Median coverage	Nuclear CG	Nuclear CHG	Nuclear CHH	Chloroplast CHH	M/P ratio
Nipponbare x Kitaake endosperm	15	43.60%	13.40%	0.80%	0.20%	2.06
Kitaake x Nipponbare endosperm	14	43.90%	11.90%	0.80%	0.20%	1.96
Nipponbare x Kitaake embryo	14	47.00%	23.00%	2.50%	0.20%	1.03
Kitaake x Nipponbare embryo	11	46.80%	22.40%	2.80%	0.20%	0.97

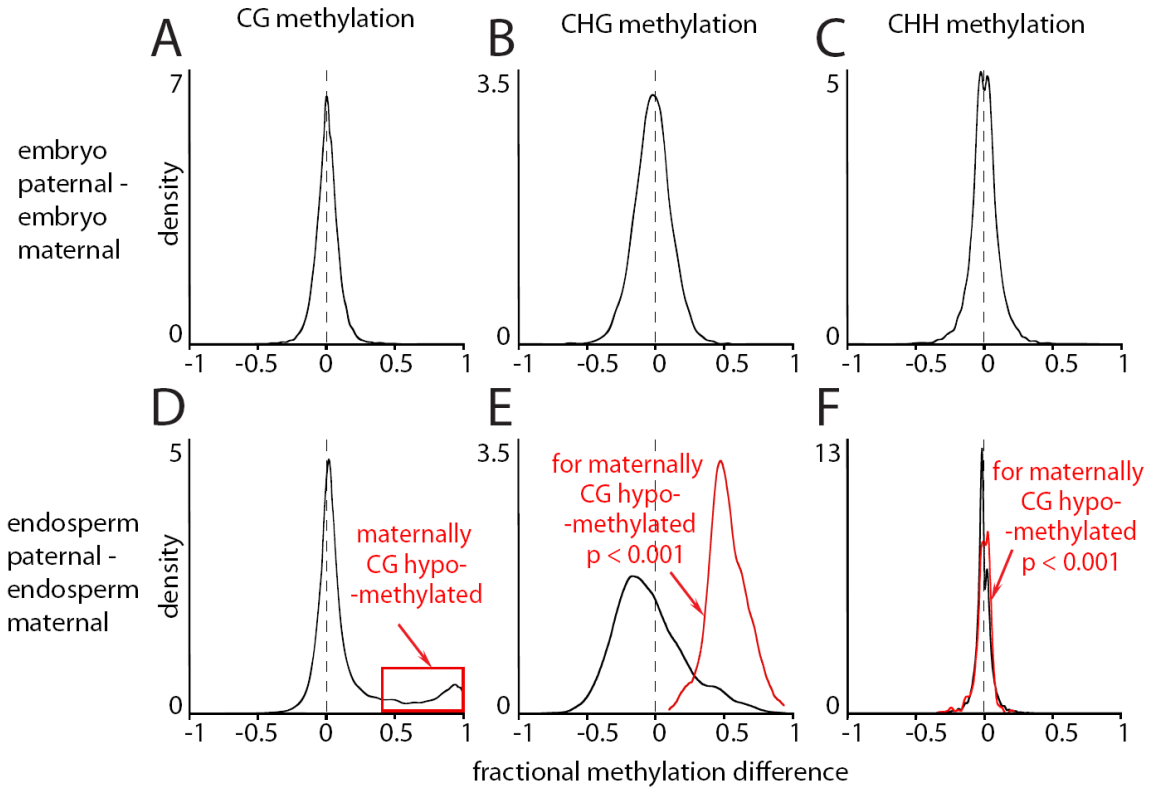


Figure 2.1 Kernel density plots showing the frequency distribution of DNA methylation differences between maternal and paternal complements in 50 bp windows across the genome for embryo (A-C) and endosperm (D-F). Red traces in (E-F) represent CHG (E) and CHH (F) methylation differences in windows that show fractional CG hypomethylation of the maternal endosperm genome greater than 0.4 (red box in D); p = p-value of a two sample Kolmogorov-Smirnov test. A shift of the main peak with respect to zero represents a global difference between maternal and paternal genomes. Shoulders at the left and right represent local hypomethylation of the paternal and maternal genomes, respectively.

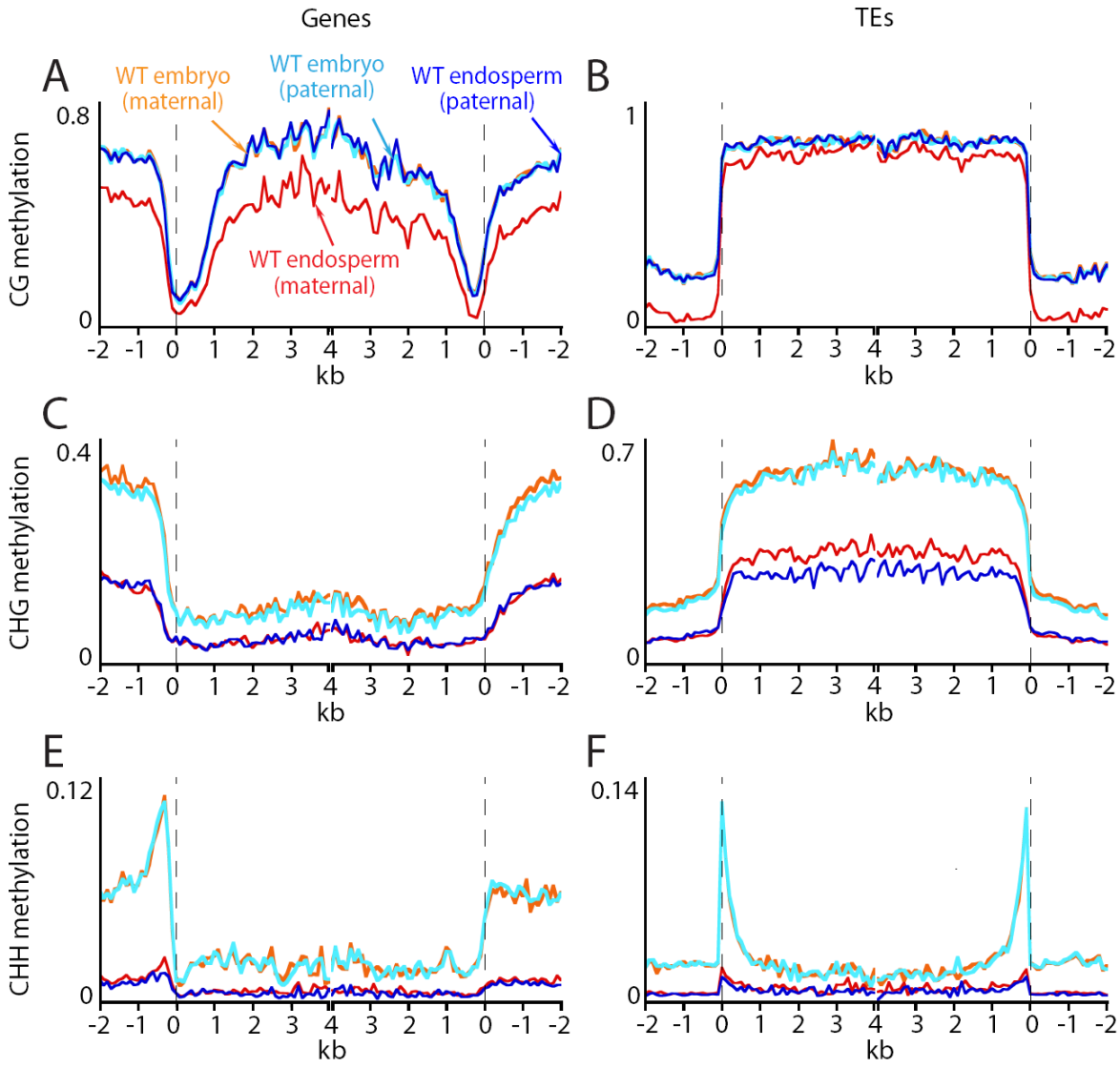


Figure 2.2 Maternal and paternal patterns of DNA methylation in rice embryo and endosperm. Genes (A, C, E) and TEs (B, D, F) were aligned at the 5' end (left panel) or the 3' end (right panel). Methylation levels within each 100-bp interval for maternal and paternal genomes were averaged between reciprocal crosses, and then plotted from 2 kb away from the annotated gene or TE (negative numbers) to 4 kb into the annotated region (positive numbers). The dashed lines at zero represent the points of alignment. CG methylation is shown in (A-B), CHG in (C-D), and CHH in (E-F).

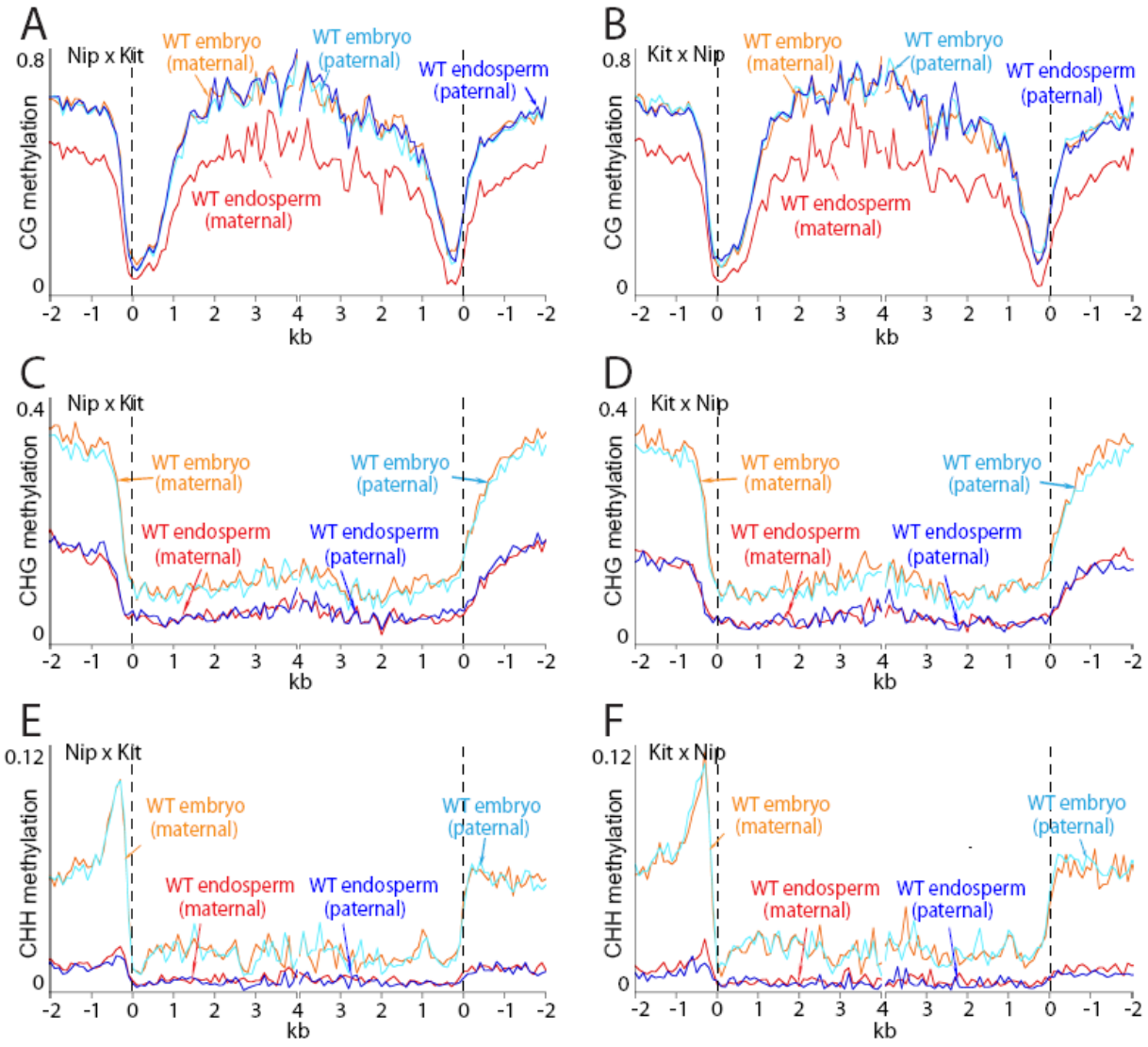


Figure 2.3 Maternal and paternal methylation of genes in endosperm and embryo of reciprocal crosses of rice. Genes were aligned at the 5' end (left panel) or the 3' end (right panel), and average methylation levels for each 100-bp interval were plotted for maternal and paternal genomes, from 2 kb away from the annotated region (negative numbers) to 4 kb into the annotated region (positive numbers). Dashed lines represent the points of alignment. CG methylation is shown in (A-B), CHG in (C-D), CHH in (E-F), with results from the Nipponbare x Kitaake cross in (A,C,E) and those from the Kitaake x Nipponbare cross in (B,D,F).

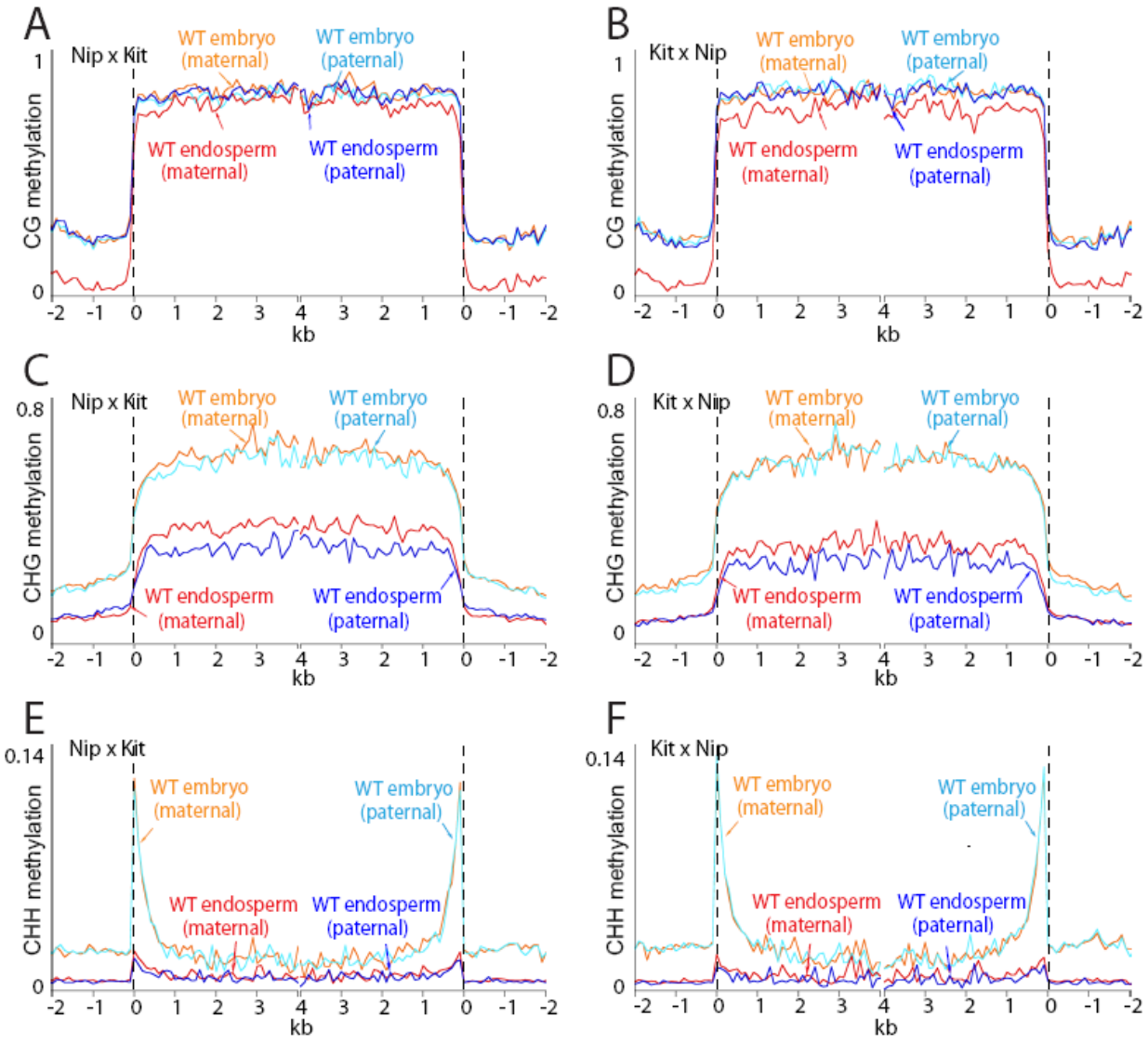


Figure 2.4 Maternal and paternal methylation of transposable element repeats in endosperm and embryo of reciprocal crosses of rice. Transposable element repeats identified by RepeatMasker were aligned at the 5' end (left panel) or the 3' end (right panel), and average methylation levels for each 100-bp interval were plotted for maternal and paternal genomes, from 2 kb away from the annotated region (negative numbers) to 4 kb into the annotated region (positive numbers). Dashed lines at zero represent the 5’ or 3’ point of alignment. CG methylation is shown in (A-B), CHG in (C-D), CHH in (E-F), with results from the Nipponbare x Kitaake cross in (A,C,E) and those from the Kitaake x Nipponbare cross in (B,D,F).

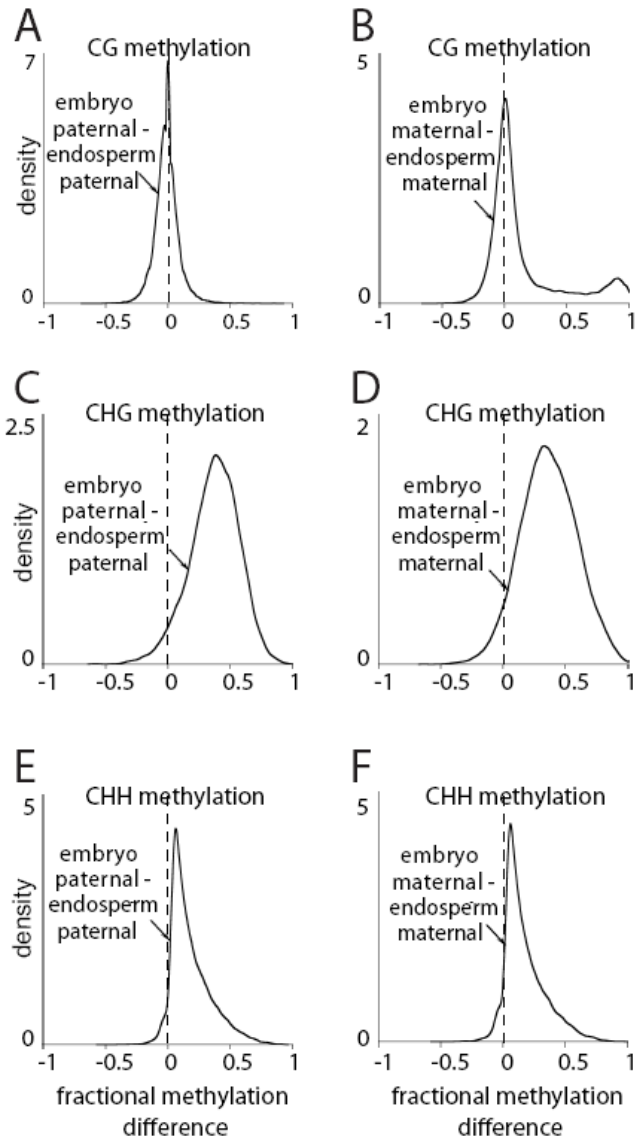


Figure 2.5 Kernel density plots showing the frequency distribution of DNA methylation differences between embryo and endosperm across the paternal genome (A,C,E) and maternal genome (B,D,F). Differences were only plotted for 50 bp windows containing at least 4 informative sequenced cytosines, and where fractional methylation of at least one of the parental genomes was greater than 0.7 in the CG context (A,B), 0.4 in the CHG context (C,D) and 0.01 in the CHH context (E,F).

Endosperm hypomethylated sites are enriched within regions of open chromatin

In *A. thaliana*, DME-mediated demethylation preferentially occurs in short TEs with euchromatic chromatin features, and is rare in gene bodies and long heterochromatic TEs (Ibarra et al., 2012). To examine the distribution of demethylated sites in rice, we identified 27,669 regions that are significantly hypomethylated at CG sites in the endosperm compared to the embryo and used them as a proxy for maternally hypomethylated sequences. These differentially methylated regions (DMRs) span 5.5% of the genome and range in size from 50 bp to 12.95 kb, with a median size of 500 bp and mean size of 737 bp.

Rice DMRs preferentially occur in genes and intergenic sequences rather than TEs (Fig. 2.6A). Among repetitive sequences, DMRs are less frequent at large Class I long terminal repeat (LTR) TEs and are more abundant in Class I non-LTR TEs, short TEs of various types, and TEs that occur within gene bodies (Fig. 2.6B-C). Although distinct from the distribution of *A. thaliana* DMRs (Ibarra et al., 2012), the depletion of rice DMRs from long TEs suggests that DME-family enzymes preferentially demethylate more accessible euchromatic sequences in rice as well as in *A. thaliana*. To test this hypothesis, we compared our methylation results with previously published DNase I hypersensitivity data (W. Zhang et al., 2012). Indeed, DNA accessibility is correlated with endosperm hypomethylation (Fig. 2.6D) – genic and intergenic regions are generally more accessible than TEs, and Class I non-LTR and shorter TEs are more accessible than longer Class I LTR and Class II TEs (Fig. 2.6D).

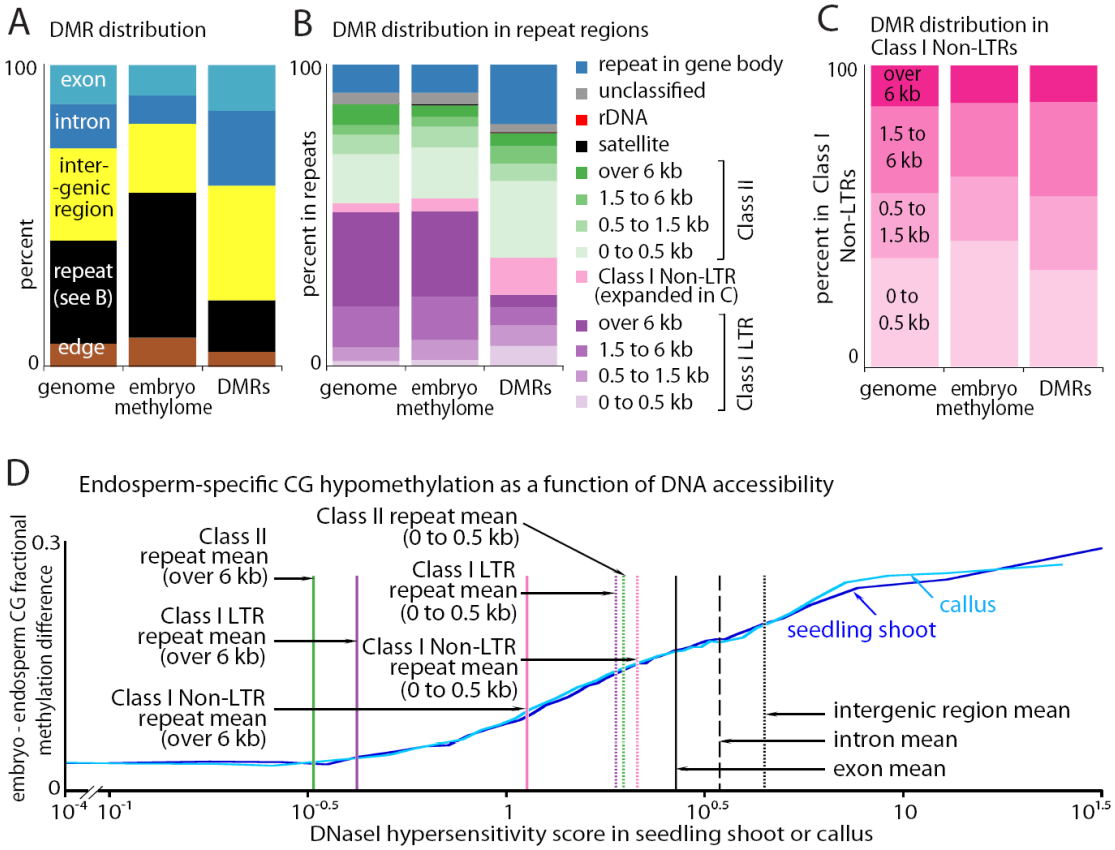


Figure 2.6 Genomic distribution of differentially methylated regions (DMRs) between embryo and endosperm. (A-C) 50-bp windows across the genome were assigned to either introns, exons, intergenic regions (no gene or repeat annotation), repeat regions (mostly TEs), or an edge category (the boundaries of gene bodies and repeats), and the number of windows that constitute the embryo methylome and the number that overlapped defined DMRs were counted. Further resolution of repeat windows (black bar in A) is shown in (B). Further resolution of Class I non-LTR sequences (pink bar in B) is shown in (C). (D) The mean CG methylation difference between embryo and endosperm was calculated for 50-bp windows of varying degrees of DNase I hypersensitivity in two tissues. The mean DNase I hypersensitivity of some sequence elements categorized in (A-C) is indicated.

Differentially methylated regions are enriched around and within imprinted genes

Because of the strong association between imprinted gene expression and endosperm hypomethylation (Bauer & Fischer, 2011; Choi et al., 2002), we examined DNA methylation at genes known to be imprinted in rice endosperm (Luo et al., 2011). Both maternally and paternally expressed genes are preferentially demethylated compared to other genes (Fig. 2.7A) and tend to have DMRs in their promoter regions (Fig. 2.7B). Maternally expressed genes are also specifically enriched for DMRs that span the transcription start site (TSS), transcription termination site (TTS), and 3' region, whereas paternally expressed genes are also specifically enriched for DMRs in their bodies (Fig. 2.7B-C and 2.8). Methylation of the TSS and TTS is correlated with gene silencing (Zemach, McDaniel, et al., 2010), consistent with the specific activation of maternally expressed endosperm genes by DNA demethylation.

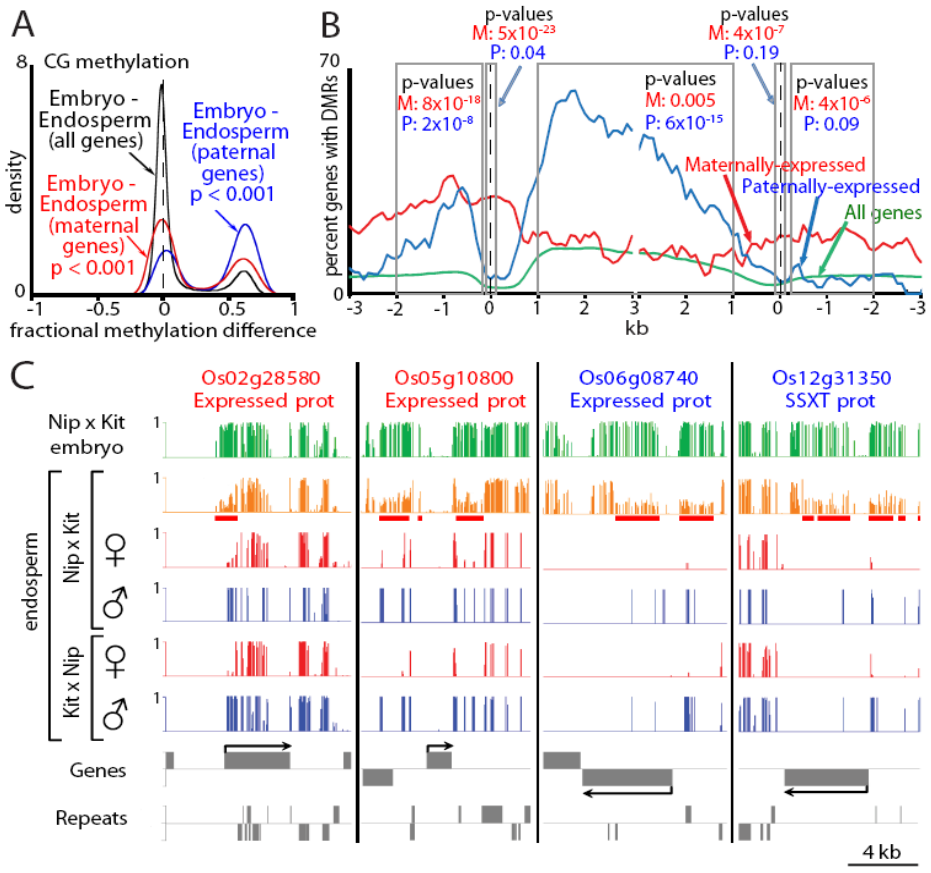


Figure 2.7 Enrichment of DMRs between embryo and endosperm at imprinted genes in rice. (A) Kernel density plots of the differences between embryo and endosperm CG methylation in 50 bp windows across the bodies of all annotated genes (black trace) and imprinted subsets (red and blue traces); p = p-value of a two sample Kolmogorov-Smirnov test. (B) Genes were aligned at the TSS and TTS, and the proportion of genes with a DMR present was plotted for 100-bp intervals within 3 kb of the alignment sites (dashed lines). At specific genic regions (grey boxes), maternally (red) and paternally (blue) expressed genes are enriched for DMRs compared to the genome average (green);); p = p-value of a Fisher's exact test. (C) Maternally (red) and paternally (blue) expressed genes enriched in DMRs. Green bars represent embryo CG methylation, orange bars represent endosperm CG methylation, and red and blue bars represent CG methylation of the maternal and paternal genomes, respectively. Identified DMRs are underlined in red.

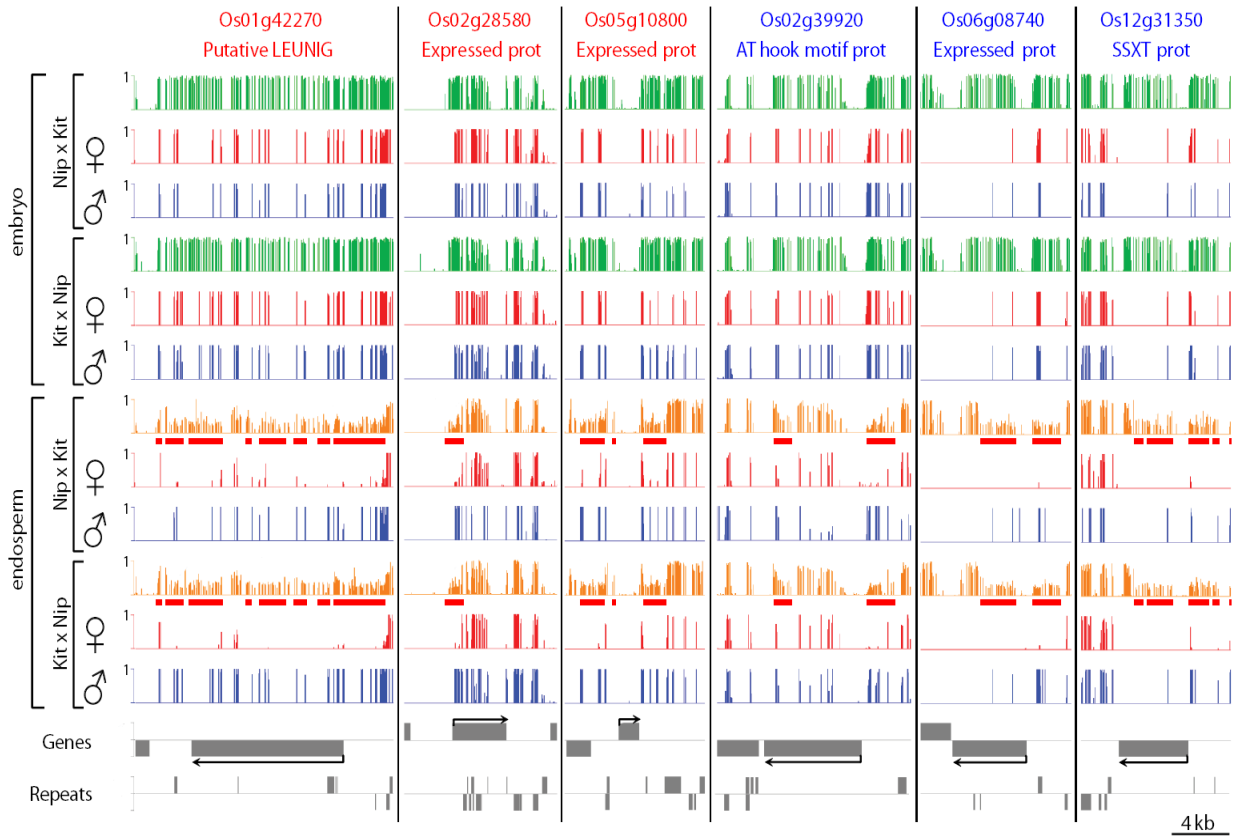


Figure 2.8 Snapshots of CG methylation in indicated rice tissues near maternally expressed (red) and paternally expressed (blue) rice imprinted genes. Green bars represent embryo methylation, orange bars represent endosperm methylation, and red and blue bars represent methylation of the maternal and paternal genomes, respectively. DMRs identified between embryo and endosperm are underlined in red.

Maternal hypomethylation patterns around imprinted genes are conserved between *Arabidopsis* and rice

DNA methylation information from *Arabidopsis thaliana* wild-type endosperm and *dme* mutant endosperm (Ibarra et al., 2012) was used to identify DMRs created by DME activity in the endosperm. We identified 9,816 endosperm DMRs and determined DMR distribution with respect to imprinted genes (Figure 2.9). The list of A. thaliana imprinted genes was obtained by combining the 114 maternally expressed genes and 9 paternally expressed genes from (Hsieh et al., 2011), 39 maternally expressed genes and 27 paternally expressed genes from (Wolff et al., 2011) and 165 maternally expressed genes and 43 paternally expressed genes from (Gehring et al., 2011). As the genome of *Arabidopsis thaliana* is generally more sparsely methylated than that of rice (Zemach, McDaniel, et al., 2010), methylated regions associated with genes in *Arabidopsis* are smaller than genic methylated regions in rice. However, we were still able to detect an enrichment of DMRs that mirrors our finding in rice: maternally expressed genes are significantly enriched for endosperm-specific DNA hypomethylation in the promoter and transcription start site, while paternally expressed genes tended to be preferentially

hypomethylated in the promoter and gene body. Thus, though the imprinting status and methylation profile of individual genes may vary between plant species, trends associated with maternally- and paternally expressed genes are consistent, further reinforcing the mechanistic similarities between monocot and dicot imprinting.

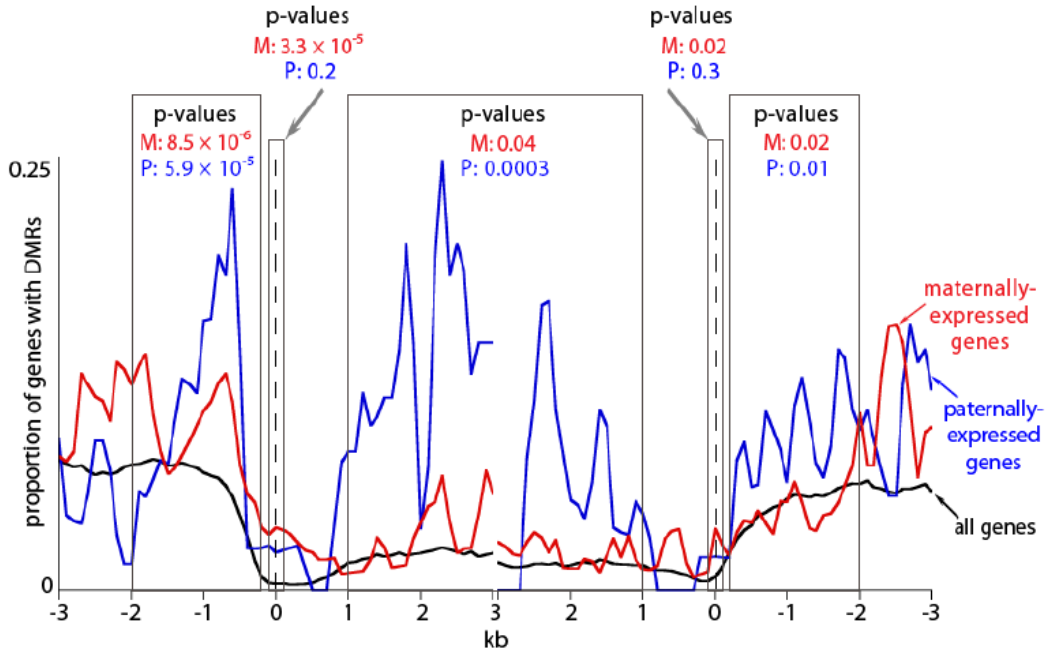


Figure 2.9 Enrichment of maternal DNA demethylation at imprinted genes in *Arabidopsis thaliana*. The distribution of differentially methylated regions between *dme* mutant and wild-type endosperm within 3 kb of each transcription terminus was plotted for maternally expressed genes (red), paternally expressed genes (blue), and the total set of annotated genes (Chen et al.). The left dashed line represents the 5' end of annotated genes and the right dashed line represents the 3' end of annotated genes. The distribution of differentially methylated regions was plotted as the proportion of genes with differentially methylated regions in 100-bp windows, with the significance of the enrichment relative to the total set at specific regions (gray boxes) indicated by a Fisher's exact test p-value.

Maternally expressed genes are mostly silenced in non-endosperm tissues while paternally expressed genes are mostly active in non-endosperm tissues

To test our theory, that the default state for many maternally expressed genes is a silenced state while that for paternally expressed genes is an active state, we compared expression of imprinted subsets to that of endosperm-expressed genes assessable for imprinting in rice leaves (Chodavarapu et al., 2012), rice shoots and roots at various developmental stages (Secco et al., 2013), and rice seedlings (T. Lu et al., 2012). Figure 2.9 shows that, for 217 paternally expressed genes and 178 maternally expressed genes described in Chapter IV, expression of the subset of maternally expressed genes in non-endosperm tissues was significantly lower than that of paternally expressed genes and the subset of endosperm-expressed genes assessable for imprinting (i.e. possessing at least one single nucleotide polymorphism in coding sequence and being expressed at over 0.15 reads per kilobase per million mapped reads in our dataset).

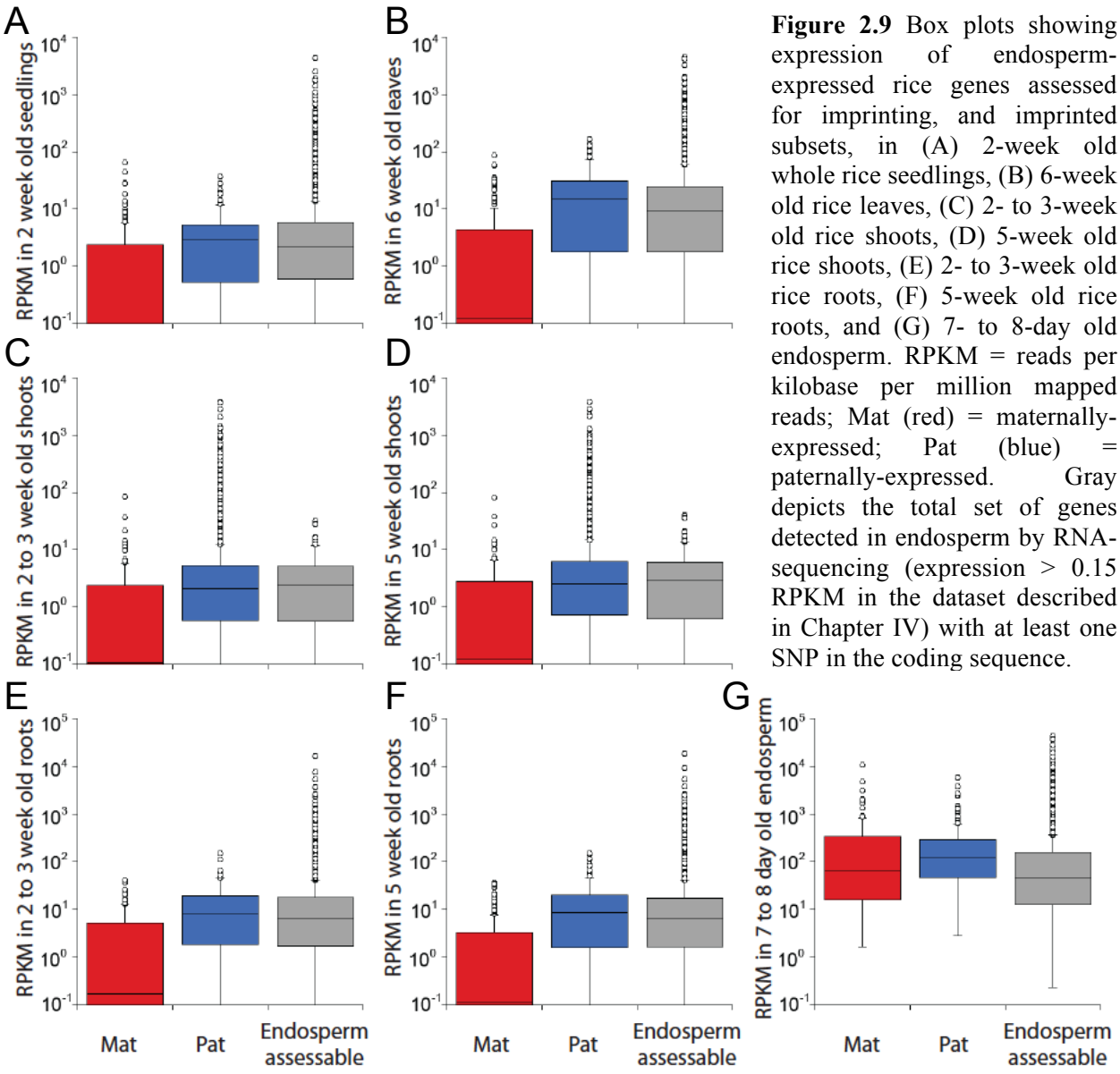


Figure 2.9 Box plots showing expression of endosperm-expressed rice genes assessed for imprinting, and imprinted subsets, in (A) 2-week old whole rice seedlings, (B) 6-week old rice leaves, (C) 2- to 3-week old rice shoots, (D) 5-week old rice shoots, (E) 2- to 3-week old rice roots, (F) 5-week old rice roots, and (G) 7- to 8-day old endosperm. RPKM = reads per kilobase per million mapped reads; Mat (red) = maternally-expressed; Pat (blue) = paternally-expressed. Gray depicts the total set of genes detected in endosperm by RNA-sequencing (expression > 0.15 RPKM in the dataset described in Chapter IV) with at least one SNP in the coding sequence.

Discussion

Our results indicate that strong site-specific CG endosperm hypomethylation is a conserved feature of flowering plants. As in *A. thaliana* (Ibarra et al., 2012), rice CG hypomethylation occurs at discrete sites with open, euchromatic features—as evidenced by DNase I sensitivity (W. Zhang et al., 2012) in Fig. 2.6D—on the maternal but not paternal chromosomes, and is associated with CHG and CHH hypomethylation at the same sites. This strongly implicates the activity of a DME-family glycosylase in the rice central cell. The rice *DME* homolog *ROS1A* (*Os01g11900*) has an expression pattern and mutant phenotype in reproductive tissues similar to *A. thaliana DME* (A. Ono et al., 2012), and is a good candidate for a functional rice DME analog. Conserved action of DME-like enzymes in monocots is also

supported by the reports of localized DNA hypomethylation in the maize central cell (Gutierrez-Marcos et al., 2006) and on the maternal genome in the endosperm (Gutierrez-Marcos et al., 2006; Waters et al., 2011).

Furthermore, *A. thaliana* and rice show similar patterns of DMR enrichment in and around maternally and paternally expressed genes, suggesting that the mechanisms that regulate imprinted expression are conserved despite the apparent lack of overlap between known imprinted genes in *A. thaliana*, rice and maize (Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2011; M. Zhang et al., 2011). Hypomethylation of maternal alleles of maternally expressed genes at the promoter and transcription start site appears to specifically activate expression of the maternal allele, while the paternal allele remains silenced. As the silencing role of DNA methylation at the transcription start site and promoter has been documented across kingdoms in plants, fungi, and animals (Gehring & Henikoff, 2007; Suzuki & Bird, 2008), this strongly suggests a default silenced state of maternally expressed genes in the meristematic lineages that form the gametophytes, followed by transcriptional activation of the maternal allele by demethylation. On the other hand, the role of gene body methylation is more nuanced and less clearly understood (Takuno & Gaut, 2013), but it anticorrelates with polycomb group activity (Weinhofer et al., 2010) and has been shown to deter polycomb-mediated silencing at polycomb-regulated genes (M. Du et al., 2014; Villar et al., 2009). Thus, hypomethylation of maternal alleles of paternally expressed genes at the promoter and gene body appears to result in polycomb-mediated silencing of the maternal allele, while the paternal allele continues to be expressed. This suggests that the default state of paternally expressed genes in the meristematic lineages that form the gametophytes is one of active expression. Data from various vegetative tissues (Figure 2.9) support this model. Our rice data are the first to corroborate findings in maize and *Arabidopsis thaliana* that implicate maternal-specific DNA demethylation as a primary imprint at both maternally and paternally expressed imprinted genes.

Materials and Methods

Nipponbare and Kitaake rice cultivars

Nipponbare and Kitaake seeds were germinated in petri dishes at 28° with light for 1 week before transplantation to field soil in an environment-controlled green house at the UC Davis Core facility. Green house conditions during the period of growth included a temperature between 27-30°C, humidity at 50%, and natural sunlight.

The two *japonica* cultivars of rice originate from Japan, with Kitaake growing in the harsher conditions of the northern Hokkaido prefecture and Nipponbare flourishing in the more temperate lower latitudes of Japan. Part of Kitaake's adaptation to the harsher climate is its fairly rapid life cycle (about 9 weeks from seed to seed), cold tolerance, and insensitivity to photoperiod, humidity, and wind agitation when flowering. Nipponbare is a high-yielding modern cultivar and does well with high fertilizer treatment; it is also grown outside of Japan, in regions with temperate climate such as Europe. Both Kitaake and Nipponbare are primarily grown to provide grain for food.

Resequencing of the Kitaake rice cultivar

Fourteen-day-old, green house-grown Kitaake seedlings were used to prepare genomic DNA using sucrose-based extraction buffer for extraction of nuclei and guanidine-based lysis buffer. gDNA was broken into smaller fragments via nebulization and ligated with adapters. Sequencing was performed using Illumina genome analyzer II at the DOE Joint Genome. Analysis was performed using maq-0.7.1 mapping and assembly software (H. Li, Ruan, & Durbin, 2008). The *Oryza sativa* ssp. *Japonica* cv. Nipponbare genome sequence MSU version 6.0 was used as reference to map Kitaake reads. Bases were called using Bustard 14.0 (80 x 1). After filtering, 169,819 single nucleotide polymorphisms (SNPs) were discovered between Kitaake and Nipponbare genomes. The sequence data have been submitted to the NCBI Sequence Read Archive under accession number SRX037797.

Isolation of rice endosperm and embryos

Reciprocal crosses were performed between Nipponbare and Kitaake cultivars of rice through manual emasculation of rice flowers followed by artificial pollination. Self-fertilized seeds from both parent varieties and F1 seeds from crosses were harvested seven to eight days after pollination. The palea and lemma were separated from the seed coat and a razor blade was used to slice open the seed coat. Milky stage endosperm was pipetted out from inside the seed coat and stored in 2% CTAB (Hexadecyltrimethylammonium bromide) for DNA extraction of genomic DNA. Embryos were isolated after the endosperm had been collected and were washed individually through vigorous agitation in 0.5 ml of 1x Phosphate Buffered Saline solution. Individually isolated F1 seeds were verified for heterozygosity with a PCR-based assay using microsatellite marker RM1 (McCouch et al., 2002; http://www.gramene.org/db/markers/marker_view?marker_id=24985466).

Bisulfite sequencing library construction and sequencing for rice

Paired-end bisulfite sequencing libraries for Illumina sequencing were constructed as described previously (T.-F. Hsieh et al., 2009) with minor modifications. In brief, about 1µg of genomic DNA was fragmented by sonication, end repaired, and ligated to custom-synthesized methylated adapters (Eurofins MWG Operon) according to the manufacturer's (Illumina) instructions for gDNA library construction. Adaptor-ligated libraries were subjected to two successive treatments of sodium bisulfite conversion using the EpiTect Bisulfite kit (Qiagen) as outlined in the manufacturer's instructions. One quarter of each bisulfite-converted library was PCR amplified using the following conditions: 2.5 U of ExTaq DNA polymerase (Takara Bio), 5 µl of 10X Extaq reaction buffer, 25 µM dNTPs, 1 µl Primer 1.1, and 1 µl Primer 2.1 (50 µl final). PCR reactions were carried out as follows: 95℃ 3 min, then 12-14 cycles of 95℃ 30 sec, 65℃ 30 sec and 72℃ 60 sec. The enriched libraries were purified twice with solid phase reverse immobilization (SPRI) method using AM-Pure beads (Beckman Coulter), prior to quantification with a Bioanalyzer (Agilent). Sequencing performed on the Illumina HiSeq 2000 platform by the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley generated 100 bp single end reads.

Rice genomic sequences and annotations

All analyses were performed with either the Nipponbare rice reference genome (MSU 6.1) or a Kitaake pseudo-genome built using the Nipponbare reference and the list of SNPs we identified in Kitaake, as described (Hsieh et al., 2011). The gene annotations we used were the MSU version 6.1 gene annotations, while repeats were annotated using RepeatMasker with the Viridiplantae Repbase database of repetitive sequenes.

In order to survey the distribution of DMRs and small RNAs within various genomic features, we assigned all 50 bp windows in the genome to one of five broad categories: exons, introns (excluding TE-derived repeat sequences that occur in gene bodies), intergenic regions, repeats (RepeatMasker-identified TE-derived repeat sequences, including those that occur in gene bodies), and regions at the boundaries of gene and repeat annotations ('edge').

Allele-specific mapping of rice reads

Reads mapped using Bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009) were sorted to either the Nipponbare reference genome or Kitaake pseudo-genome as described (Hsieh et al., 2011), and DNA methylation of cytosines within sorted reads was calculated as described (T.-F. Hsieh et al., 2009; Zemach, McDaniel, et al., 2010).

Generation of kernel density plots for rice DNA methylation data

Kernel density plots of paternal-maternal differences in fractional methylation were generated using 50 bp windows where at least 4 cytosines were informative and the fractional methylation of at least one of the parental genomes was greater than the defined threshold for a particular context. Thresholds were 0.7 for the CG context, 0.4 for the CHG context, and 0.01 for the CHH context. Kernel density plots of embryo-endosperm differences in fractional CG methylation were generated using 50 bp windows where at least 19 cytosines contributed to the overall methylation value in each cross, and in which at least one of the tissues had a methylation value greater than 0.7 and the methylation values of reciprocal crosses were within 0.1 of each other.

Definition of differentially methylated regions (DMRs) in rice

CG fractional methylation in 50-bp windows (each window averages methylation on both strands) was compared between embryo and endosperm. Windows with embryo fractional methylation at least 0.2 greater than that in endosperm and a Fisher's exact test p-value less than 0.05 were merged if they occurred within 300 bp. Merged DMRs were retained if the fractional methylation in embryo across the DMR was at least 0.3 greater than that in endosperm and the Fisher's exact test p-value was less than 10^{-5}.

Definition of differentially methylated regions (DMRs) in *Arabidopsis thaliana*

Fractional CG methylation in 50-bp windows across the genome was compared between *dme* endosperm and wild-type endosperm. Windows with a fractional CG methylation difference of at least 0.3 between *dme* endosperm and wild-type endosperm (Fisher's exact test p-value < 0.001) were merged to generate larger differentially methylated regions (DMRs) if they occurred

within 300 bp. DMRs were retained for further analysis if the fractional CG methylation across the whole DMR was 0.3 greater in *dme* endosperm than in wild-type endosperm (Fisher's exact test p-value < 10^{-10}), and if the DMR was at least 100 bp.

Expression of endosperm-expressed genes in non-endosperm tissues

Publically available RNA sequencing reads from rice leaves (Chodavarapu et al., 2012), rice shoots and roots at various developmental stages (Secco et al., 2013), and rice seedlings (T. Lu et al., 2012) was aligned to cDNA scaffolds of rice genes using Bowtie (Langmead et al., 2009). For all annotated genes, one isoform was chosen to be a representative cDNA scaffold for each gene, and the expression of the isoform was assessed by calculating the number of reads mapped per kilobase of isoform sequence, per million reads that mapped to the set of cDNA scaffolds. Where biological replicate RNA sequencing datasets were available for the same tissue-type, the RPKM counts of individual libraries were averaged together to produce a final estimate for gene expression level.

Chapter III

Imprinted small RNAs in rice endosperm sometimes overlap imprinted genes and are also associated with localized hypomethylation of the maternal genome

The following chapter has been published as part of a peer reviewed article in the Proceedings of the National Academy of Sciences, USA:

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D (2013). Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proceedings of the National Academy of Sciences, USA, 110* (19): 7934-7939. doi: 10.1073/pnas.1306164110

Contributions:
Pamela Ronald, Robert Fischer, and Daniel Zilberman conceived and designed the experiment, obtained NSF funding, and provided discussion and advice at various stages. Randy Ruan peformed rice crosses and assisted in dissecting and harvesting rice endosperm and embryo. Toshiro Nishimura wrote and modified Perl Scripts for part of the data analysis. Manoj and Rita Sharma coordinated the resequencing of the Kitaake rice cultivar and identification of single nucleotide polymorphisms between Kitaake and Nipponbare rice cultivars with staff at the Joint Genome Institute. I contributed to experimental design, dissected and harvested rice endosperm and embryo, performed all molecular biology experiments, wrote custom Python scripts for the small RNA pipeline and performed all data analysis.

Sequencing data are deposited in GEO with accession number GSE44898.

Introduction

Small RNAs (sRNAs) in the *A. thaliana* seed appear to be strongly imprinted, with the majority of identified 24-nt sRNAs being derived from maternal sequences (Mosher et al., 2009). These sRNAs appear to accumulate in the endosperm and mediate gene expression (J. Lu et al., 2012; Mosher et al., 2009). However, accumulation of maternally derived sRNA is unaffected by mutations in DNA- and histone-modifying enzymes that are known to regulate imprinted gene expression, including the DEMETER DNA glycosylase (Mosher et al., 2011), leaving the mechanism by which they are generated unresolved. Rice, a monocot that diverged from *A. thaliana* roughly 150 million years ago (Chaw et al., 2004), possesses marked decreases in CG methylation at specific sites throughout the genome that correlate with endosperm-specific gene expression (Zemach, Kim, et al., 2010) and imprinted gene expression. However, we currently

do not know whether sRNAs in the rice seed are generally seed-specific or maternally biased. Here, we show that, unlike in *A. thaliana*, imprinted 24-nt sRNAs in the endosperm originate from both parental genomes and are associated with demethylated regions, closely resembling the behavior of protein coding genes and suggesting that sRNA accumulation may be regulated differently in the persistent endosperm of rice and the more ephemeral endosperm of *A. thaliana*.

Results

Rice endosperm exhibits a unique sRNA expression pattern

To examine sRNA expression in rice seeds, we sequenced endosperm and embryo sRNA libraries from Nipponbare and Kitaake reciprocal crosses, and also libraries from several control tissues (Table 3.1). Embryo tissues, on average, produced a larger fraction of 24-nt small RNAs than other tissues, though significant variation was observed between biological replicates for control seedling tissues. The endosperm did not appear to produce a larger fraction of 24-nt small RNAs than control seedling tissues.

Table 3.1 Summary statistics for small RNA libraries.

Small RNA library (17 to 30 nt reads)	Total number of sequenced reads	Size class	Number of reads in size class	Percent size class forms of total	Reads that aligned	Percent of reads that aligned	Reads that sorted to Nippon--bare (Nip)	Percent aligned reads that sorted to Nip	Reads that sorted to Kitaake (Kit)	Percent aligned reads that sorted to Kit
Nipponbare x Kitaake embryo	27464694	17 nt	129414	0.47%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	273478	1.00%	258084	94.37%	592	0.23%	582	0.23%
		19 nt	427837	1.56%	395800	92.51%	1206	0.30%	1088	0.27%
		20 nt	812923	2.96%	761710	93.70%	2185	0.29%	2058	0.27%
		21 nt	1768610	6.44%	1685549	95.30%	4891	0.29%	4552	0.27%
		22 nt	1635290	5.95%	1532946	93.74%	8399	0.55%	8007	0.52%
		23 nt	4939312	17.98%	4575730	92.64%	29146	0.64%	28082	0.61%
		24 nt	16362199	59.58%	15723275	96.10%	108094	0.69%	102911	0.65%
		25 nt	644139	2.35%	594195	92.25%	4308	0.73%	4136	0.70%
		26 nt	87572	0.32%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare x Kitaake endosperm	28662079	17 nt	305621	1.07%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	716815	2.50%	647968	90.40%	771	0.12%	490	0.08%
		19 nt	1110177	3.87%	978584	88.15%	1508	0.15%	776	0.08%
		20 nt	1753968	6.12%	1536625	87.61%	2673	0.17%	1388	0.09%
		21 nt	3712458	12.95%	3290785	88.64%	6230	0.19%	3106	0.09%
		22 nt	1993446	6.95%	1744638	87.52%	5856	0.34%	2793	0.16%
		23 nt	4216408	14.71%	3670482	87.05%	16074	0.44%	8003	0.22%
		24 nt	13000441	45.36%	12051207	92.70%	60844	0.50%	29017	0.24%
		25 nt	696522	2.43%	585307	84.03%	2130	0.36%	1006	0.17%
		26 nt	255299	0.89%	N/A	N/A	N/A	N/A	N/A	N/A

Small RNA library (17 to 30 nt reads)	Total number of sequenced reads	Size class	Number of reads in size class	Percent size class forms of total	Reads that aligned	Percent of reads that aligned	Reads that sorted to Nippon--bare (Nip)	Percent aligned reads that sorted to Nip	Reads that sorted to Kitaake (Kit)	Percent aligned reads that sorted to Kit
Kitaake x Nipponbare embryo	27516451	17 nt	146411	0.53%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	309452	1.12%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	499800	1.82%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	919315	3.34%	830218	90.31%	2626	0.32%	2370	0.29%
		21 nt	1833247	6.66%	1685489	91.94%	5400	0.32%	5054	0.30%
		22 nt	1834159	6.67%	1653747	90.16%	9263	0.56%	8957	0.54%
		23 nt	5175278	18.81%	4610070	89.08%	30632	0.66%	28721	0.62%
		24 nt	15482066	56.26%	14445116	93.30%	102569	0.71%	95867	0.66%
		25 nt	622457	2.26%	N/A	N/A	N/A	N/A	N/A	N/A
		26 nt	90041	0.33%	N/A	N/A	N/A	N/A	N/A	N/A
Kitaake x Nipponbare endosperm	24820055	17 nt	451562	1.82%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	802859	3.23%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	1124504	4.53%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	1754929	7.07%	1566075	89.24%	1002	0.06%	1680	0.11%
		21 nt	3876532	15.62%	3578740	92.32%	2571	0.07%	4306	0.12%
		22 nt	1608618	6.48%	1427286	88.73%	1835	0.13%	3251	0.23%
		23 nt	2964792	11.95%	2609187	88.01%	5568	0.21%	9949	0.38%
		24 nt	8295660	33.42%	7879939	94.99%	19727	0.25%	35015	0.44%
		25 nt	633790	2.55%	N/A	N/A	N/A	N/A	N/A	N/A
		26 nt	328311	1.32%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare root 1	19860042	17 nt	149714	0.75%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	307646	1.55%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	427080	2.15%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	687116	3.46%	586570	85.37%	857	0.15%	68	0.01%
		21 nt	1465296	7.38%	1242660	84.81%	1985	0.16%	101	0.01%
		22 nt	1046634	5.27%	598612	57.19%	3241	0.54%	215	0.04%
		23 nt	3706780	18.66%	955598	25.78%	10272	1.07%	187	0.02%
		24 nt	11171671	56.25%	2109735	18.88%	25982	1.23%	348	0.02%
		25 nt	401851	2.02%	170244	42.36%	1366	0.80%	30	0.02%
		26 nt	90853	0.46%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare root 2	16745583	17 nt	450959	2.69%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	641198	3.83%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	735813	4.39%	N/A	N/A	N/A	N/A	N/A	N/A

Small RNA library (17 to 30 nt reads)	Total number of sequenced reads	Size class	Number of reads in size class	Percent size class forms of total	Reads that aligned	Percent of reads that aligned	Reads that sorted to Nippon--bare (Nip)	Percent aligned reads that sorted to Nip	Reads that sorted to Kitaake (Kit)	Percent aligned reads that sorted to Kit
Nipponbare root 2	16745583	20 nt	1112615	6.64%	1004560	90.29%	1839	0.18%	135	0.01%
		21 nt	2190449	13.08%	1933410	88.27%	4091	0.21%	152	0.01%
		22 nt	1216375	7.26%	1003457	82.50%	5773	0.58%	238	0.02%
		23 nt	2503809	14.95%	1657108	66.18%	17059	1.03%	282	0.02%
		24 nt	6401722	38.23%	3721715	58.14%	46085	1.24%	536	0.01%
		25 nt	331942	1.98%	246259	74.19%	2217	0.90%	65	0.03%
		26 nt	96787	0.58%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare shoot 1	22494930	17 nt	311214	1.38%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	778825	3.46%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	985602	4.38%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	2214117	9.84%	2117334	95.63%	2861	0.14%	201	0.01%
		21 nt	3403549	15.13%	3215557	94.48%	6117	0.19%	290	0.01%
		22 nt	2015088	8.96%	1903508	94.46%	10344	0.54%	538	0.03%
		23 nt	3739419	16.62%	3622253	96.87%	37040	1.02%	646	0.02%
		24 nt	7438216	33.07%	7217283	97.03%	88750	1.23%	1217	0.02%
		25 nt	580745	2.58%	545693	93.96%	3964	0.73%	120	0.02%
		26 nt	255012	1.13%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare shoot 2	23209105	17 nt	204800	0.88%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	401636	1.73%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	514833	2.22%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	1193703	5.14%	1111495	93.11%	1998	0.18%	119	0.01%
		21 nt	2339237	10.08%	2185117	93.41%	4990	0.23%	173	0.01%
		22 nt	1493714	6.44%	1279247	85.64%	8013	0.63%	305	0.02%
		23 nt	4195220	18.08%	3118796	74.34%	32926	1.06%	479	0.02%
		24 nt	11416378	49.19%	7658653	67.08%	94389	1.23%	1046	0.01%
		25 nt	604492	2.60%	471341	77.97%	4436	0.94%	98	0.02%
		26 nt	182339	0.79%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare endosperm	20224400	17 nt	211754	1.05%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	349319	1.73%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	581271	2.87%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	1182255	5.85%	1054108	89.16%	2208	0.21%	136	0.01%
		21 nt	3380386	16.71%	3111122	92.03%	6130	0.20%	248	0.01%
		22 nt	1384959	6.85%	1252834	90.46%	5974	0.48%	220	0.02%

Small RNA library (17 to 30 nt reads)	Total number of sequenced reads	Size class	Number of reads in size class	Percent size class forms of total	Reads that aligned	Percent of reads that aligned	Reads that sorted to Nippon--bare (Nip)	Percent aligned reads that sorted to Nip	Reads that sorted to Kitaake (Kit)	Percent aligned reads that sorted to Kit
Nipponbare endosperm	20224400	23 nt	2925221	14.46%	2758375	94.30%	20119	0.73%	389	0.01%
		24 nt	7895746	39.04%	7540637	95.50%	62448	0.83%	1248	0.02%
		25 nt	468864	2.32%	418582	89.28%	2078	0.50%	44	0.01%
		26 nt	195080	0.96%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare embryo	24768963	17 nt	179895	0.73%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	231703	0.94%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	395552	1.60%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	840437	3.39%	730734	86.95%	4655	0.64%	206	0.03%
		21 nt	1718121	6.94%	1573063	91.56%	10002	0.64%	274	0.02%
		22 nt	2026273	8.18%	1873059	92.44%	19248	1.03%	609	0.03%
		23 nt	5513136	22.26%	5167945	93.74%	61782	1.20%	981	0.02%
		24 nt	12228650	49.37%	11404575	93.26%	144933	1.27%	2350	0.02%
		25 nt	446346	1.80%	389730	87.32%	5202	1.33%	129	0.03%
		26 nt	74826	0.30%	N/A	N/A	N/A	N/A	N/A	N/A
Kitaake endosperm	21412573	17 nt	462443	2.16%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	517477	2.42%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	671490	3.14%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	985272	4.60%	675571	68.57%	231	0.03%	1947	0.29%
		21 nt	2079907	9.71%	1690950	81.30%	225	0.01%	4432	0.26%
		22 nt	1232091	5.75%	841588	68.31%	272	0.03%	3950	0.47%
		23 nt	3153345	14.73%	1975528	62.65%	706	0.04%	12394	0.63%
		24 nt	9420465	44.00%	5749375	61.03%	1771	0.03%	40677	0.71%
		25 nt	470342	2.20%	273318	58.11%	67	0.02%	1349	0.49%
		26 nt	167772	0.78%	N/A	N/A	N/A	N/A	N/A	N/A
Kitaake embryo	25216437	17 nt	206551	0.82%	N/A	N/A	N/A	N/A	N/A	N/A
		18 nt	308277	1.22%	N/A	N/A	N/A	N/A	N/A	N/A
		19 nt	461342	1.83%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	791042	3.14%	666594	84.27%	127	0.02%	3230	0.48%
		21 nt	1562688	6.20%	1400028	89.59%	156	0.01%	7059	0.50%
		22 nt	1596593	6.33%	1406733	88.11%	175	0.01%	13665	0.97%
		23 nt	5175303	20.52%	4595154	88.79%	465	0.01%	54638	1.19%
		24 nt	13614884	53.99%	11826479	86.86%	1268	0.01%	150824	1.28%
		25 nt	619458	2.46%	512617	82.75%	105	0.02%	6167	1.20%
		26 nt	134960	0.54%	N/A	N/A	N/A	N/A	N/A	N/A

However, unlike in embryo, seedling root and seedling shoot, 24-nt sRNAs in the endosperm predominantly map to genic and intergenic regions rather than transposable elements (Fig. 3.1A), though embryos show a somewhat intermediate pattern (Fig. 3.1A). Furthermore, the endosperm 24-nt sRNA reads that do map to transposable elements (TEs) indicate a different pattern of TE-associated sRNA expression compared to other tissues (Fig. 3.1B-C). The majority of TE-associated sRNA production in embryo, shoot and root occurs within Class II TEs shorter than 500 bp (Fig 3.1B). This group is largely comprised of miniature inverted-repeat transposable elements (MITEs), which preferentially occur near genes in euchromatic regions and are the predominant target of CHH methylation in rice tissues (C. Lu et al., 2012; Zemach, Kim, et al., 2010). In contrast, 24-nt production in endosperm is much more evenly distributed across different TE classes and sizes (Fig. 3.1B), though Class I TEs over 6 kb are still underrepresented compared to their genomic abundance (Fig. 3.1B-C).

The endosperm is also distinguished by strong expression of 24-nt sRNA from a modest number of loci (Fig. 3.1D). We termed the genomic regions of relatively high sRNA production in the endosperm 'siren' loci (small-interfering RNA in endosperm). Siren loci are dispersed intermittently across all twelve chromosomes, accounting for 74% of 24-nt sRNA in Nip x Kit endosperm and 64% of 24-nt sRNA in Kit x Nip endosperm, despite cumulatively spanning only 0.44% of the genome. Siren loci also appear to preferentially locate to genes, intergenic regions and small TEs within the genome (rightmost bars in Fig. 3.1A-C).

Imprinting of endosperm 24-nt small RNA shares similarities with that of genes

The stringency associated with accurately mapping short reads to parental genomes, along with the sparse distribution of SNPs within regions of sRNA production, resulted in a very small fraction of the total reads from F1 seeds being informative for detection of parental biases. For example, of the 12,991 endosperm sRNA-producing loci we identified, only 125 possessed a total of at least 15 parentally-sorted reads in each of the reciprocal crosses. Nevertheless, we were able to assay the general trends of 24-nt sRNA parental expression in embryo and endosperm.

We did not detect significant parental bias in the embryo (Fig. 3.1E), but did find large deviations from the expected 2:1 maternal:paternal ratio in the endosperm (Fig. 3.1E). Of the 125 informative loci, 15 had a significant maternal bias ($p < 0.001$; Fisher's exact test) and 16 had a paternal bias in both reciprocal crosses. Only four loci were unambiguously biallelic. Like imprinted genes, imprinted 24-nt sRNA-producing loci are enriched in DMRs (Fig. 3.1F). These results are in contrast to the strong tendency toward maternal sRNA expression described in *A. thaliana* seeds (Mosher et al., 2009) that is unaffected by DNA methylation (Mosher et al., 2011).

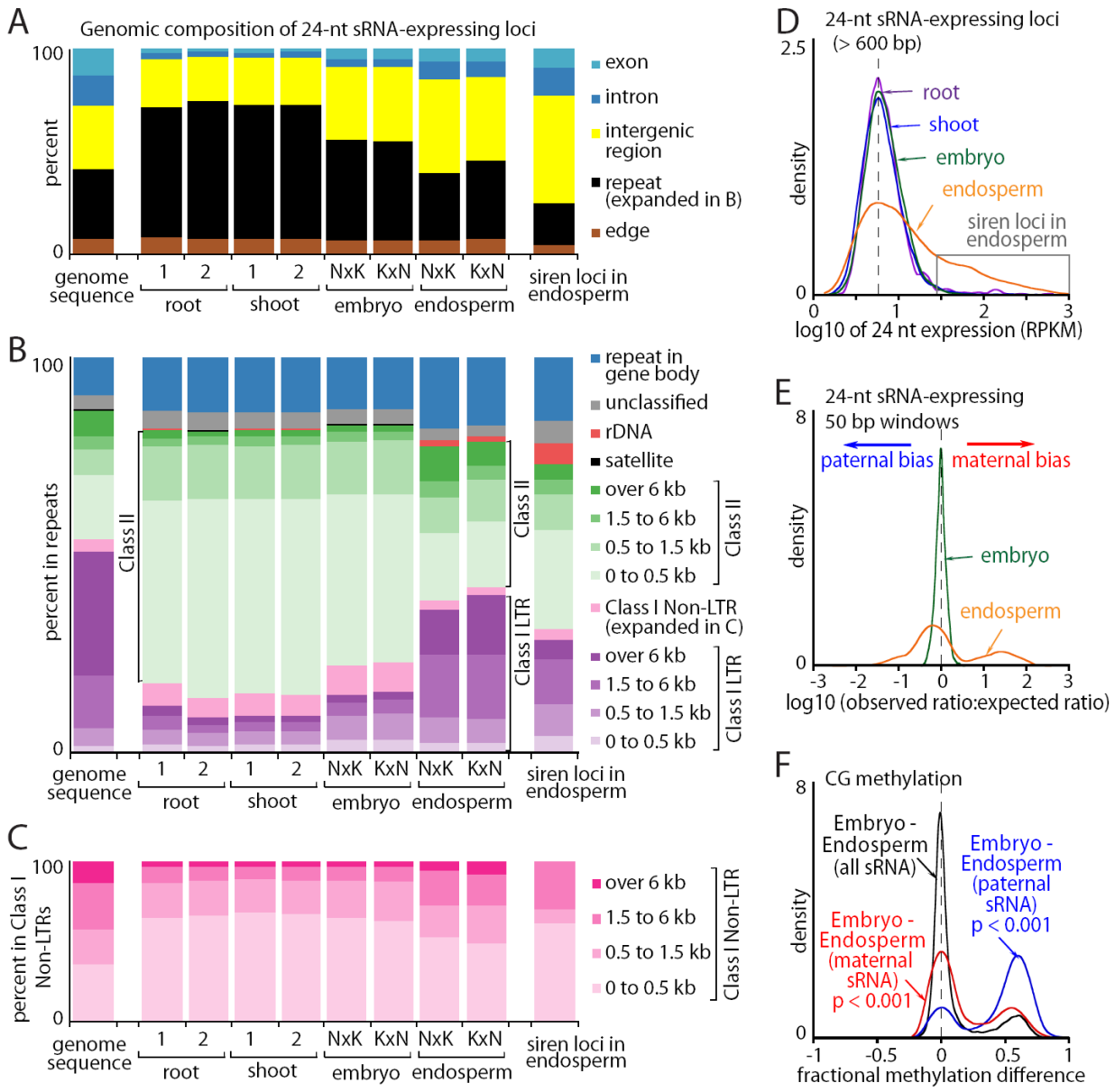


Figure 3.1 24-nt small RNA-expression in the endosperm, embryo, seedling root, and seedling shoot. (A-C) Relative abundance of 24-nt sRNA expresson across the genome, subdivided as in Fig. 2.6A-C. (D) Kernel density plot of 24-nt sRNA expression intensity in endosperm, embryo, root, and shoot at loci longer than 600 bp (length threshold imposed to exclude microRNA loci). (E) Kernel density plots depicting the prevalence and direction of 24-nt sRNA parental bias. Deviations from zero (dashed line) indicate either maternal bias (deviation to the right) or paternal bias (deviation to the left). (F) Kernel density plots of the differences between embryo and endosperm CG methylation measured in 50-bp windows across all sRNA-producing loci (black trace) and imprinted subsets (red and blue traces); p = p-value of a two sample Kolmogorov-Smirnov test.

Some imprinted genic loci are associated with imprinted 24-nt small RNA

Of the 31 imprinted sRNA loci, four overlap known imprinted genes, with sRNA and mRNA expression always occurring from opposite parental alleles (Fig. 3.2 and Table 3.2). Of the remaining 27 imprinted sRNA loci, 18 overlap genes of unknown imprinted status, suggesting that a substantial proportion of imprinted sRNA may correspond to imprinted genes. Notably, the maternally expressed gene *Os02g29230* (middle panel in Fig. 3.2), which overlaps DMRs and paternally expressed imprinted sRNAs, encodes a DME homolog named *ROS1B* (A. Ono et al., 2012), suggesting the possibility of a complex, multilayered regulatory relationship between DNA demethylation and sRNA pathways.

Table 3.2 Regions of overlap between imprinted 24-nt small RNA-expressing loci and imprinted genes.

Chr	sRNA locus start	sRNA locus end	sRNA locus expression bias	Gene locus ID	Gene locus start	Gene locus end	Gene locus expression bias	Functional annotation of gene
1	40538301	40544848	maternal	Os01g70060	40539921	40553708	paternal	RNA Pol-II binding domain
2	17344451	17357050	paternal	Os02g29230	17348970	17357069	maternal	DEMETER homolog
6	19595101	19605500	paternal	Os06g33690	19603058	19605119	maternal	CAPIP1
7	4701351	4709900	paternal	Os07g09020	4701864	4711518	maternal	AGO14

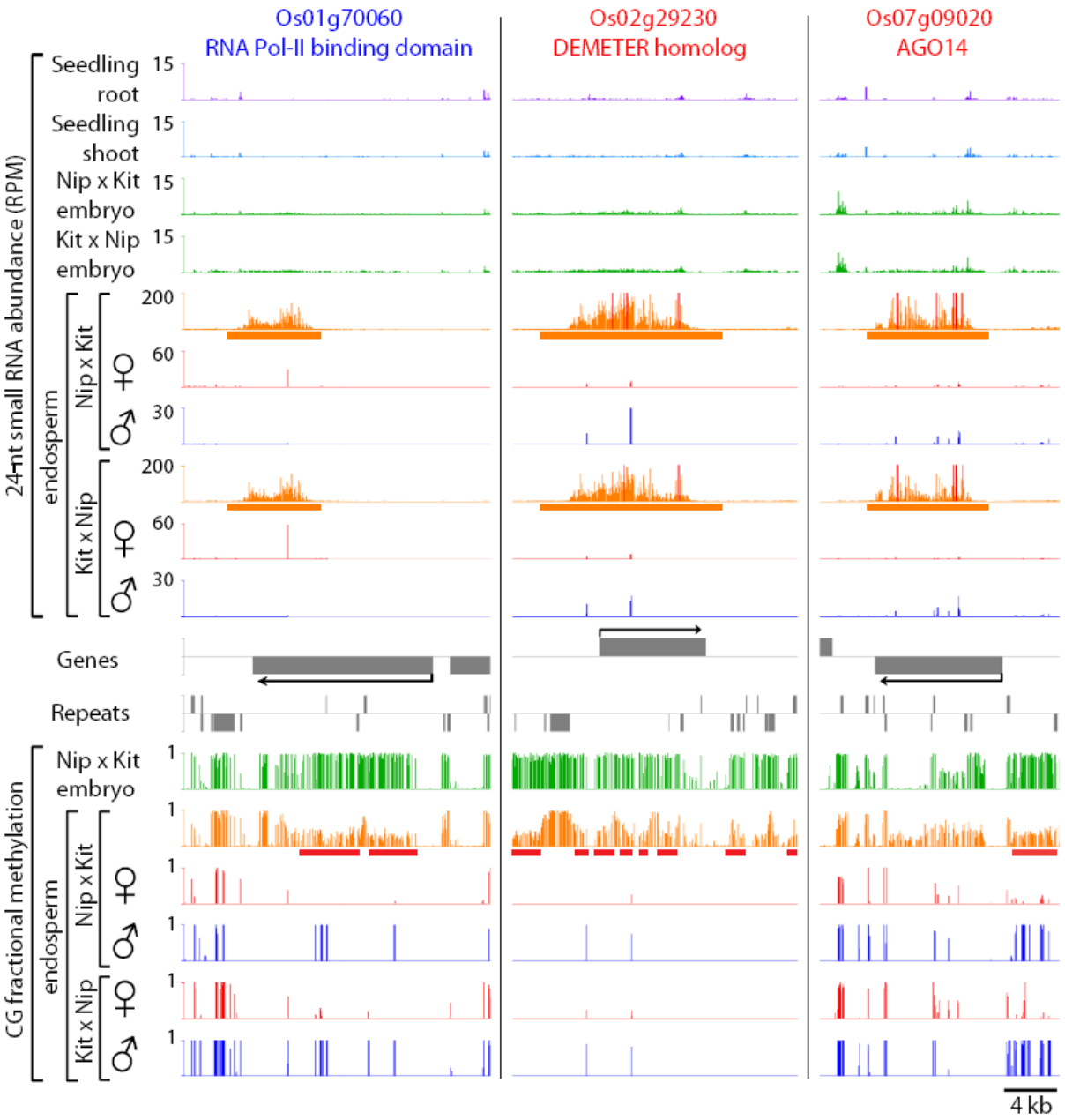


Figure 3.2 Snapshots of 24-nt sRNA abundance in reads per million (RPM) and CG methylation in rice embryo and endosperm, around paternally expressed (blue) and maternally expressed (red) imprinted genes. At positions where SNPs resolved reads between parental genomes, 24-nt sRNA abundance and DNA methylation of the maternal and paternal genome are represented by red bars and dark blue bars, respectively. Siren loci identified in our analysis are underlined in orange, DMRs identified between embryo and endosperm are underlined in red. Note the different scales for endosperm sRNA compared to the other tissues. Scales for sRNA sorted to maternal and paternal genomes reflect the 2:1 ratio of maternally- to paternally-inherited chromosomes in the endosperm.

Discussion

We were the first to show paternally biased expression of small RNAs in a plant system and to visualize the genomic distribution of endosperm small RNAs. We also show that rice sRNA-producing loci behave like genes with respect to imprinting, particularly in terms of the direction of imprinting biases and correlation with sites of CG hypomethylation. Paternal expression of endosperm small RNAs associated with maternal DMRs may be regulated by the same epigenetic mechanisms that have been proposed to explain paternal mRNA expression linked to maternal DMRs (Bauer & Fischer, 2011). Taken together with the likely conservation of imprinting mechanisms in flowering plants, this suggests that conventionally imprinted sRNA loci may also exist in *A. thaliana*, but have not been detected because of the large amount of maternally encoded sRNA that is not regulated by known imprinting mechanisms (Mosher et al.). It is possible that this accumulation indeed reflects a novel imprinting pathway that does not operate in rice. Alternatively, the maternal bias seen in *A. thaliana* may have been due to deposition of sRNAs from surrounding maternal tissue such as the seed coat, as the ratio of seed coat tissue to endosperm is greater in *A. thaliana* seeds than in the much larger rice seeds used in our experiments.

We and others have proposed that a major function of the demethylation that occurs in central and vegetative cells is to produce mobile sRNAs that are transported into the egg and sperm cells to reinforce the methylation and silencing of TEs (Bauer & Fischer, 2011; Ibarra et al., 2012; Schoft et al., 2011; Slotkin et al., 2009). Generation of the sRNA signal may be facilitated by the relaxation of chromatin in the companion cells (Schoft et al., 2011; Slotkin et al., 2009), and is supported by unusual sRNA patterns in the *A. thaliana* pollen (Slotkin et al., 2009). Our finding that the sRNA population of rice endosperm spans a wider range of genomic sequences, including expression from many TE types that are underrepresented in other tissues, is consistent with this idea. Because the sRNA patterns in the relatively mature rice embryos we analyzed are distinct from those in endosperm, our data also strongly argue that the bulk of the hypothesized sRNA exchange must occur during early development, either between the gametes and their companion cells prior to fertilization, and/or between the nascent embryo and endosperm.

Materials and Methods

Isolation of endosperm and embryos from Nipponbare and Kitaake rice cultivars

Rice seeds were grown, reciprocal crosses were performed, and seeds were harvested as described in Chapter II. Milky stage endosperm was pipetted out from inside the seed coat of seven- to eight-day old seeds and stored in TRIzol reagent (Ambion) for preparing small RNA libraries. Embryos were isolated after the endosperm had been collected and were washed individually through vigorous agitation in 0.5 ml of 1x Phosphate Buffered Saline solution. Individually isolated F1 seeds were verified for heterozygosity with a PCR-based assay using microsatellite marker RM1 (McCouch et al., 2002; http://www.gramene.org/db/markers/marker_view?marker_id=24985466)

Preparation of rice seedling tissue

Nipponbare rice seeds were grown in sterile flask culture with Gamborg's B-5 medium with sucrose (Caisson Laboratories). At day 21 after germination, seedlings were rinsed in sterile water, and root and shoot tissue were separated and snap-frozen in liquid nitrogen for future RNA extraction with TRIzol.

Small RNA library construction and sequencing

sRNA libraries were constructed as described (Couvillion et al., 2009). In brief, total RNA was prepared from snap-frozen rice seedling root and shoot tissue, and snap-frozen embryos and endosperm using TRIzol reagent (Ambion). 10 µg of total RNA was loaded on a 15% polyacrylamide, 7M urea gel, and small RNA in the 17-30 nt size range were excised. 3' miRNA cloning linker 1 (IDT) was ligated to the gel-excised small RNA using truncated T4 RNA ligase 2 (NEB), and ligation products were purified on a 15% polyacrylamide, 7M urea gel. The 5' Illumina RNA linker (5'-rArCrArCrUrCrUrUrUrCrCrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU-3') was added using T4 RNA ligase (NEB) and ligation products were again purified on a 15% polyacrylamide, 7M urea gel. Purified ligated small RNAs were reverse transcribed using SuperScriptIII (Invitrogen) and PCR-amplified with Phusion High-Fidelity DNA Polymerase (NEB) for 25 cycles (Forward primer: 5'-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACG-3', Reverse primer: 5'-CAAGCAGAAGACGGCATACGATTGATGGTGCCTACAG-3'). PCR products were gel-purified on a 2% agarose gel to obtain fragments between 100 to 120 bp in length for sequencing in either 36 or 50 bp reads on the Illumina HiSeq 2000 (crossed F1 embryo and endosperm) and GA-II (self-fertilized embryo, endosperm, seedling root and seedling shoot) platforms.

Rice genomic sequences and annotations

All analyses were performed with either the Nipponbare rice reference genome (MSU 6.1) or a Kitaake pseudo-genome built using the Nipponbare reference and the list of SNPs we identified in Kitaake, as described in Chapter II and (Hsieh et al., 2011). The gene annotations we used were the MSU version 6.1 gene annotations, while repeats were annotated using RepeatMasker with the Viridiplantae Repbase database of repetitive sequences. In order to survey the distribution of small RNAs within various genomic features, we assigned all 50 bp windows in the genome to one of five broad categories: exons, introns (excluding TE-derived repeat sequences that occur in gene bodies), intergenic regions, repeats (RepeatMasker-identified TE-derived repeat sequences, including those that occur in gene bodies), and regions at the boundaries of gene and repeat annotations ('edge').

Allele-specific mapping of reads

Adapter sequences were trimmed from the reads and resultant trimmed reads were sorted into size classes from 17 nt to 30 nt using custom Python scripts. Bowtie (Langmead et al., 2009) was used to independently align each size class to both the Nipponbare genome and Kitaake pseudo-genome, and custom Perl scripts (Hsieh et al., 2011) were used to sort reads to one or the other parental genome. Reads were assigned to a parental genome only in instances

where a read aligned to both genomes in the same position, with zero mismatches to one and a single mismatch to the other. Abundances of total and parent-specific small RNA across the genome were calculated in reads per million (RPM).

Identification of sRNA-producing loci

Tissue-specific loci of 24-nt small RNA expression were defined using two biological replicate libraries of Nipponbare (Nip) seedling shoot, two biological replicate libraries of Nip seedling root, embryo libraries from reciprocal crosses (one Nip x Kitaake library and one Kitaake x Nip library) and endosperm libraries from reciprocal crosses (using the same seeds as for embryo). Loci of 24-nt small RNA expression were defined by blocking together 50 bp windows with over 5 read counts in both libraries from the same tissue type, when the windows were within 500 bp of each other.

Detection of imprinting of defined sRNA-producing loci

Where data from reciprocal crosses was available and could be sorted with single nucleotide polymorphisms, the number of maternally-sorted and paternally-sorted 24-nt reads that mapped within these loci was recorded. Loci with 15 or more informative reads were evaluated for significant deviations from the expected maternal-paternal ratio using Fisher's exact test. Imprinted loci were only identified in the endosperm; maternally expressed loci were defined as loci with a significant bias ($p < 0.001$) and maternal-paternal ratio greater than 3.3 in both crosses, while paternally expressed loci were defined as loci with a significant bias and maternal-paternal ratio less than 1 in both crosses. Loci that were not defined as maternally or paternally expressed were either defined to be biallelic on the basis of having a maternal-paternal ratio between 1.6 and 2.5 in both crosses, or assigned to an "unclear" category of loci that showed variety-specific bias or were parentally-biased but did not pass our significance criteria.

Chapter IV

Imprinted gene expression in rice is largely conserved over short evolutionary time

The following chapter shall soon be submitted for publication as part of a peer reviewed article.

Contributions:
Pamela Ronald, Robert Fischer, and Daniel Zilberman conceived and designed the experiment, obtained NSF funding, and provided discussion and advice at various stages. Randy Ruan performed rice crosses and assisted in dissecting and harvesting rice endosperm and embryo. Toshiro Nishimura wrote and modified Perl Scripts for data. Manoj and Rita Sharma, in conjunction with staff at the Joint Genome Institute, performed the resequencing of the Kitaake and IR64 rice cultivars and identification of single nucleotide polymorphisms between the cultivars and the Nipponbare genome reference. I contributed to experimental design, dissected and harvested rice endosperm and embryo, performed all molecular biology experiments, wrote custom Python scripts for data analysis, and performed all data analysis.

Introduction

Although advances in high-throughput sequencing technology have allowed the detection of imprinted genes at a genome-wide scale, an understanding of the evolutionary forces driving imprinting in plants remains elusive. The functions of most identified imprinted plant genes are poorly characterized or unknown at this stage (Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2011; Wolff et al., 2011; M. Zhang et al., 2011), though analyses testing enrichment of Gene Ontology terms in candidate imprinted genes in both *Arabidopsis thaliana* (Pignatta et al., 2014) and maize (Waters et al., 2013) find paternally expressed imprinted genes significantly enriched for regulatory functions, especially those relating to transcription. On the other hand, maternally expressed genes showed only a weak enrichment in functions related to transcriptional regulation in *Arabidopsis* (Pignatta et al., 2014) but were significantly enriched for functions related to transcriptional regulation and development in maize (Waters et al., 2013).

Also, it is unclear whether imprinting in plants evolves as rapidly as is expected in contexts where biotic conflict is experienced (Brockhurst et al., 2014). In the case of imprinting, this conflict could arise between maternal and paternal parents as the fitness of the male parent is increased if his offspring thrive at the cost of other offspring borne by the female parent, while the fitness of the maternal parent is increased if she is able to have as many successful offspring as possible (David Haig & Westoby, 1989). These evolutionary pressures result in paternal expression of genes that promote nutrient acquisition in the offspring and maternal expression of genes that restrict nutrient allocation to any one offspring. Not all imprinting might be driven by such parental conflict, but in cases where it is, it is likely that regulatory features or the coding sequence of a gene may evolve at a faster rate. It is interesting to note that many imprinted genes in *Arabidopsis* belong to gene families with multiple paralogs, possibly due to their enhanced potential to rapidly evolve new functions and expression patterns (Qiu, Liu, & Adams, 2014; Yoshida & Kawabe, 2013).

Although the low degree of overlap among three independent genome-wide lists of *Arabidopsis thaliana* imprinted genes (Gehring et al., 2011; Hsieh et al., 2011; Wolff et al., 2011) suggested the possibility that imprinting had evolved quickly across short evolutionary timescales of a few thousand years during the divergence of the different strains used in the three studies, a recent study comparing imprinting in three different *Arabidopsis thaliana* strains found that while 110 genes showed conserved imprinting in all assessable strains (beginning with at least two of the three) only 12 genes showed divergence in imprinting among the three strains (Pignatta et al., 2014). This suggests that the imprinting status of the vast majority of genes (90%) are conserved across strains in *Arabidopsis thaliana*.

Similarly, two independent genome-wide lists of maize imprinted genes (Waters et al., 2011; M. Zhang et al., 2011) only showed a 50% overlap, but a study of the conservation of imprinted gene expression across four maize varieties revealed conservation of imprinting status at 88% of imprinted genes (Waters et al., 2013). Thus, it appears that once variables related to plant growth condition, developmental stage, presence of informative single nucleotide polymorphisms (SNPs) and data analysis techniques are controlled for, both *Arabidopsis* and maize exhibit a trend towards imprinting conservation across short evolutionary timescales.

The first genome-wide survey of imprinted expression in rice identified 262 candidate imprinted loci in the endosperm but showed poor overlap between imprinted genes in rice and *Arabidopsis thaliana* (Luo et al., 2011), while only 13% of genes with syntenic orthologs in maize were imprinted in both species (Waters et al., 2013). In order to test the degree of imprinting conservation over short evolutionary time scales in rice, we chose four cultivars of the rice species *Oryza sativa* L. that span both the indica and japonica subspecies and were previously shown to be of importance in the rice community (Coffman & Hargrove, 1989; Wu et al., 2005; H. W. Zhang et al., 2012). We generated four sets of reciprocal crosses between Kitaake, Nipponbare, 93-11, and IR64 rice cultivars, and detected imprinted expression using single nucleotide polymorphisms (SNPs) between the rice cultivars used in each cross.

Results

Resequencing of the IR64 cultivar

We sequenced the genome of the IR64 cultivar of the indica subspecies of rice, achieving 50-fold depth of coverage. IR64 is the most widely grown indica rice in South and Southeast Asia because of its wide adaptability to different field environments, high yield potential, disease and pest resistance, and consumer appeal (Wu et al., 2005). This has made it a good candidate for genetic studies aimed at understanding rice development. In order to compare expression in IR64 to expression in other rice varieties, we identified single nucleotide polymorphisms (SNPs) with reference to MSU version 7.0 genome sequence of the Nipponbare japonica rice cultivar (Ouyang et al., 2007; International Rice Genome Sequencing Project, 2005). In short, mapped IR64 reads covered 90% of the Nipponbare genome and we identified 1,704,329 SNPs between Nipponbare and IR64.

Divergence of rice varieties as deduced by single nucleotide polymorphisms (SNPs)

The complex breeding history of rice cultivars (Coffman & Hargrove, 1989; H. W. Zhang et al., 2012) makes it difficult to accurately estimate how long ago any two rice cultivars would have begun diverging. It is likely that different loci were inherited from different ancestral rice varieties so that some loci may share a more recent last common ancestor while others share a more distant last common ancestor. However, we can estimate the degree of divergence at individual loci and across the whole genome on average by comparing the number of single nucleotide polymorphisms (SNPs) identified. A list of SNPs between Kitaake and Nipponbare was generated by us as previously described in Chapter II, while a list of SNPs between 93-11 and Nipponbare was obtained by combining two independently published lists (Huang et al., 2009; D. Wang, Xia, Li, Hou, & Yu, 2012) and eliminating SNPs that caused mismapping of either Nipponbare or 93-11 bisulfite-sequencing reads.

We find that the genomes of the two japonica cultivars, Nipponbare and Kitaake, are the most closely related to each other, while the genomes of the two indica cultivars, 93-11 and IR64, contain three times the number of SNPs seen between Nipponbare and Kitaake (Table 4.1). As expected, the number of SNPs between each indica cultivar and Nipponbare was much greater than that seen within each subspecies, being about 10 times that between the two japonica cultivars and three times that between the indica cultivars.

Table 4.1 Number of SNPs identified between cultivars based on resequencing and mapping to the MSU version 7.0 genome sequence of the Nipponbare japonica rice cultivar. Numbers in brackets indicate the percentage of the genome that possesses SNPs.

Rice variety	Kitaake (Kit) japonica cultivar	93-11 indica cultivar	IR64 indica cultivar
Nipponbare (Nip) japonica cultivar	167,335 (0.05%)	1,672,658 (0.45%)	1,704,329 (0.46%)
93-11 indica cultivar	N/A	N/A	509,253 (0.14%)

Detection of imprinted gene expression in reciprocal crosses

We performed four sets of reciprocal crosses in order to evaluate imprinted expression in the four chosen rice cultivars: one set of intra-japonica crosses (i.e. Nipponbare × Kitaake and Kitaake × Nipponbare), one set of intra-indica crosses (i.e. IR64 × 93-11 and 93-11 × IR64), and two sets of inter-subspecies crosses (Nipponbare × IR64 and IR64 × Nipponbare, and Nipponbare × 93-11and 93-11 × Nipponbare). Individual seeds from each cross were dissected and harvested at day 7 to 8 after pollination and genotyped to verify that they were indeed the result of the intended cross-pollination. Strand-specific RNA-sequencing (RNA-seq) libraries generated from endosperm-extracted RNA were subjected to duplex-specific nuclease (DSN) treatment in order to enhance transcript detection. Trials comparing sequencing results from DSN-treated libraries before and after treatment showed that maternal and paternal biases, both at individual loci and on average across the transcriptome, were not affected by DSN treatment (Figure 4.1A,B).

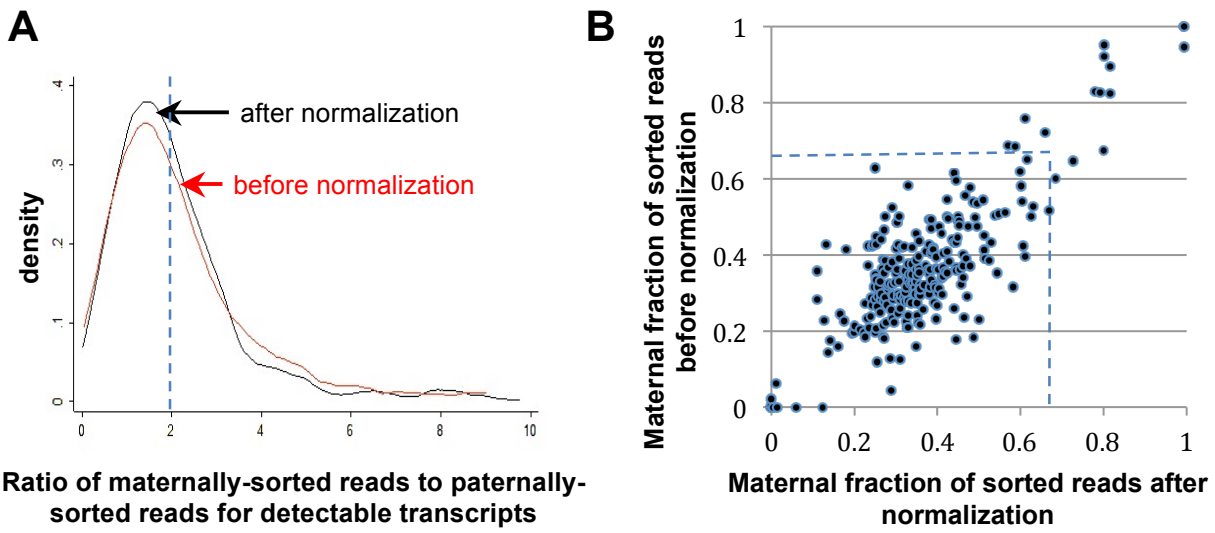


Figure 4.1 Duplex-specific nuclease treatment of RNA-seq libraries does not alter the ratio of maternally- and paternally-sorted reads. (A) Kernel density plots show similar distributions of the ratio of maternally-sorted reads to paternally-sorted reads for detectable transcripts of the Kitaake × Nipponbare endosperm RNA-seq library before and after normalization. (B) Scatter plot comparing the fraction of sorted reads that are maternally-sorted in individual genes before and after normalization. Dashed lines indicate the expected ratio or fraction of maternally-sorted reads.

After mapping DSN-treated RNA-seq reads to cDNA models for each annotated gene in the MSU version 7.0 genome sequence of the Nipponbare japonica rice cultivar (Ouyang et al., 2007; International Rice Genome Sequencing Project, 2005) with the Bowtie read aligner, we were able to detect expression of about 35 to 42% of annotated rice genes, depending on the cross (Figure 4.2A). While a significant number of detectable transcripts either lacked informative SNPs to sort reads to parental genomes or did not possess enough sorted reads to make conclusions about their parental biases, we could evaluate parental biases in 20% of annotated genes (i.e. 47 to 55% of detectable transcripts) for the inter-subspecies crosses and 2 to 6% of annotated genes (5 to 17% of detectable transcripts) for intra-subspecies crosses. However, about a third of detectable transcripts with a sufficient number of sorted reads exhibited variety-specific expression in reciprocal crosses (i.e. expression was consistently biased towards a particular cultivar rather than one of the parental genomes), and we were unable to make clear calls about imprinted expression of either parental cultivar in this situation. Of the remaining genes, which are the genes truly assessable for imprinting, parental biases varied along a continuum, as previously noted in rice (Luo et al., 2011). We found that roughly 10% of expressed genes appear to be imprinted in a moderate, strong, or complete manner, as defined by a maternally-derived fraction of sorted reads that deviates significantly (Fisher's exact test, $p<0.01$) from the expected ratio of 0.67 (Figure 4.2B).

Detection of conservation and divergence of imprinting among rice cultivars

Having identified hundreds of genes that appear to display parental biases in a manner unrelated to the varieties involved in the cross (Figure 4.2C), we inferred the imprinting status of

parental varieties using data from all available reciprocal cross data. A schematic pipeline detailing the method used to detect conserved imprinting status is shown in Figure 4.3, while the additional steps used to verify imprinting divergence are shown in Figure 4.4 and 4.5.

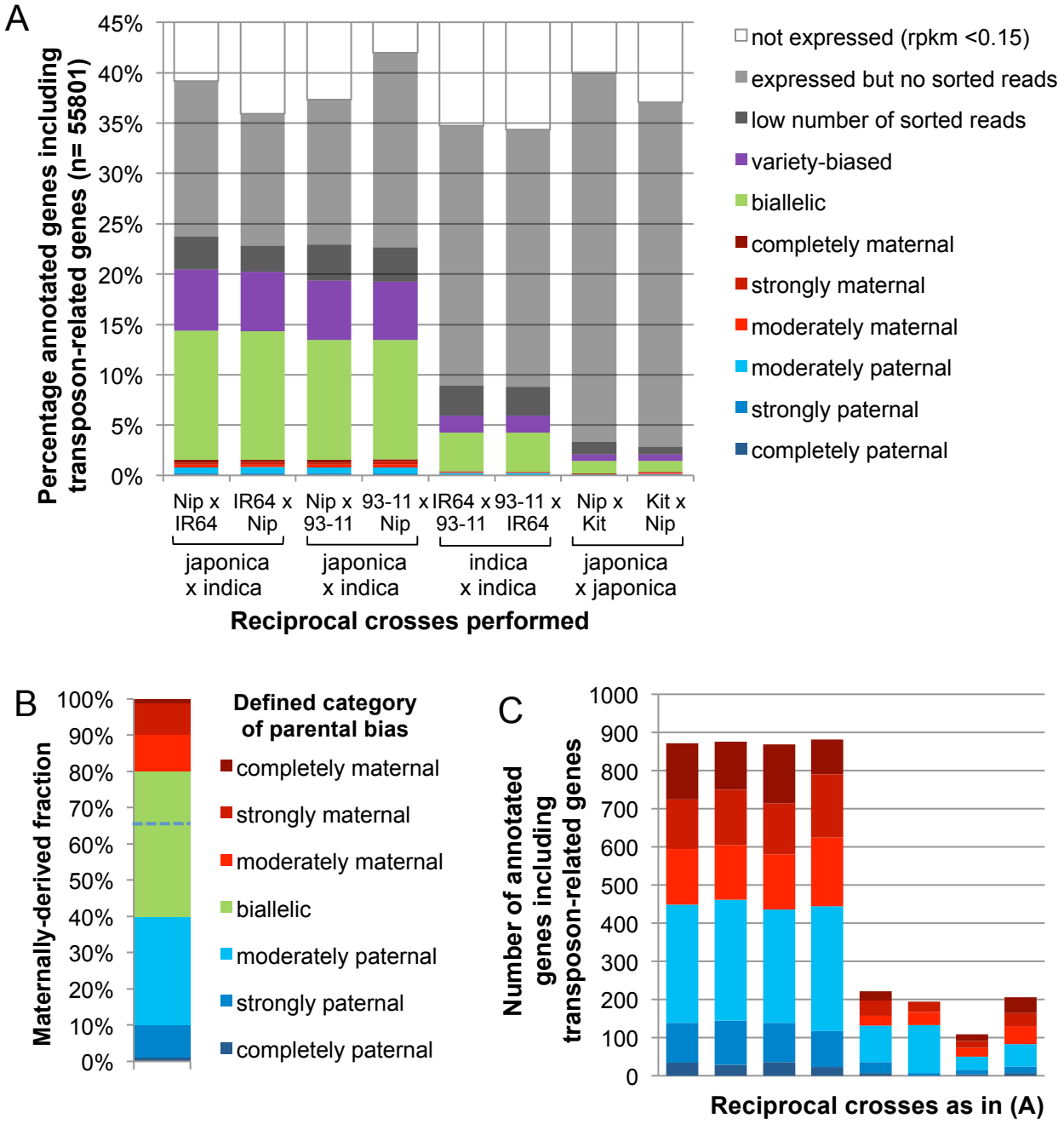


Figure 4.2 Detection of imprinted genes in reciprocal crosses. (A) Genes with detectable expression were sorted into various categories based on their assessability for parental biases and, in case they were assessable, the degree of parental bias. (B) If the parental bias of a gene deviated significantly (Fisher's exact test, $p < 0.01$) from the expected ratio for endosperm (dashed line), it was classified to be either moderately, strongly (over 90% uniparental) or completely (over 99% uniparental) parentally-biased. (C) Hundreds of genes exhibited significant parental biases.

16034 genes (including transposon-related genes) with informative SNPs for imprinting detection

1) Filter genes with low coverage out by **eliminating** reciprocal cross data where:
   a) At least one cross has less than 15 informative reads
   b) The maternal fraction of at least one cross passes imprinting thresholds but is not statistically supported (Fisher's exact test, $p < 0.01$), due to a low number of reads
   c) Both reciprocal crosses are assigned conflicting imprinting calls but the difference in maternal fraction is insignificant (Fisher's exact test, $p \geq 0.01$), due to low coverage
   d) At least one cross has gene expression below 0.15 reads per kilobase per million mapped reads (RPKM)

2) Filter variety-specific biases out by **eliminating** reciprocal cross data where:
   a) Crosses in the reciprocal set have statistically significant opposing parental biases (Fisher's exact test, $p < 0.01$), due to the dominance of a particular variety
   b) Crosses in the reciprocal set show a four-fold difference in gene expression, due to variety-specific variation in expression level
   c) One cross in the reciprocal set is called to be 'imprinted' and the other called 'biallelic', but each deviates from the expected maternal fraction and in opposing directions. This situation is detected by checking that the average maternal fraction of such a set of reciprocal crosses is between 0.6 and 0.7
   d) Both crosses in the reciprocal set are called to be 'biallelic' but the difference in maternal fraction is greater than 0.2

9894 genes with good coverage and informative SNPs for imprinting detection

For the 9894 genes with information, imprinting was assessed for individual varieties based on the combined data from all reciprocal crosses:

- 7316 genes were biallelic across all assessable varieties (i.e. 2 or more varieties)
- 251 genes only had ‘paternally expressed’ imprinting calls across all assessable varieties*
- 237 genes only had ‘maternally expressed’ imprinting calls across all assessable varieties**
- 658 genes had both ‘maternally expressed’ and ‘biallelic’ calls across assessable varieties***
- 733 genes had both ‘paternally expressed’ and ‘biallelic’ calls across assessable varieties***
- 699 genes had conflicting imprinting calls for one or more varieties**

*Of these 251 genes, 196 could be verified to be imprinting in at least three cultivars.

**Closer inspection of the 237 genes that only had ‘maternally expressed’ imprinting calls across all assessable varieties revealed some of them were likely due to contamination of endosperm tissue with maternally-derived seed coat tissue. Filtering to eliminate genes potentially affected by contamination reduced the list to 189 genes, of which 127 genes were verified to be imprinted in at least 3 cultivars.

***The latter three groups provided a list of 2090 candidate examples of imprinting divergence but, as shown in Figure 4.4, most of these 2090 genes were falsely categorized due to the presence of mismapped reads, erroneous SNPs, weak variety-specific biases, poor complexity of RNA-sequencing reads at particular loci, and multiple transcripts arising from the same locus.

Figure 4.3 Schematic illustration of pipeline used to detect conserved imprinting across rice cultivars.

Contamination of endosperm samples with seed coat RNA

We observed that nine nuclear genes encoding photosynthesis-related proteins, such as members of photosystem I and II and carbon fixing enzymes, appear to be maternally expressed in rice endosperm, even though rice endosperm is a non-photosynthetic tissue (Burkhardt et al., 1997). Inspection of these genes, revealed that they exhibit 50- to 1815-fold higher expression in non-endosperm tissues such as rice leaves (Chodavarapu et al., 2012), rice shoots and roots at various developmental stages (Secco et al., 2013), and rice seedlings (T. Lu et al., 2012). Additionally, these genes do not possess the MEG-type endosperm DNA hypomethylation profile previously described (Ibarra et al., 2012; Rodrigues et al., 2013). In the absence of an available dataset describing RNA expression in the rice seed coat, we assumed that the nine photosynthesis related genes were contaminants from the seed coat, which is the only photosynthetic tissue in close contact with the endosperm. It is possible that the rupturing of the seed coat, a procedure required for the dissection of endosperm tissue, released seed coat RNA into the endosperm sample.

If there is indeed contamination of endosperm tissue with seed coat RNA, it is likely that there are genes other than the nine photosynthesis-related genes that appear to be imprinted and maternally expressed but are instead expressed highly in the seed coat and possibly not expressed or imprinted in the endosperm. In order to filter our list of maternally expressed genes for such contamination events, we used previously published data from photosynthetic rice tissues that are likely to have a higher overlap of gene expression with the seed coat than with endosperm. We filtered genes where expression measured in reads per million per kilobase (RPKM) in the endosperm was less than 5% of that in 6 week old rice leaves (Chodavarapu et al., 2012), rice shoots at 2 to 3 weeks and 5 weeks after germination (Secco et al., 2013), and 14-day old rice seedlings (T. Lu et al., 2012). This additionally eliminated various enzymes, transporters and a few transcriptional regulators, reducing the list of 237 genes that appeared to be maternally expressed in all assessable varieties to 189. Further analysis of genes that appeared to be divergent cases of imprinting revealed many false positive calls that were instead conserved cases of imprinting, so the final list of maternally expressed genes is presented later.

4 to 11% of rice imprinted genes diverge across cultivars due to a loss or gain of imprint

Figure 4.4 describes how many of the initial apparent divergences in imprinting identified by our pipeline (Figure 4.3) proved to be technical artefacts upon careful manual inspection. From a list of 421 genes with enough information to estimate imprinting divergence, we chose 174 strong candidates for manual verification. For this analysis, RNA-seq reads were mapped to the genome using the TopHat read aligner, which accounts for splice junctions and is thus able to map reads to novel gene isoforms and unannotated loci, reducing instances of read mismapping and improving estimates of transcript structure and the parental bias of individual exons. Out of the initial 174 genes, we find that 55 genes do not possess sufficient informative reads for imprinting assessment, 51 genes appear maternally expressed in all assessable varieties, 21 genes appear paternally expressed in all assessable varieties, 14 genes appear biallelic in all assessable varieties, 8 appear to be contaminants from surrounding maternal tissues, and 7 appear to be prone to developmental resetting as previously described in rice endosperm (Luo et al., 2011). Of the remaining 18 genes, 9 appear to vary between maternally expressed and biallelic states across

cultivars while the other 9 appear to vary between paternally expressed and biallelic states across cultivars (Table 4.2).

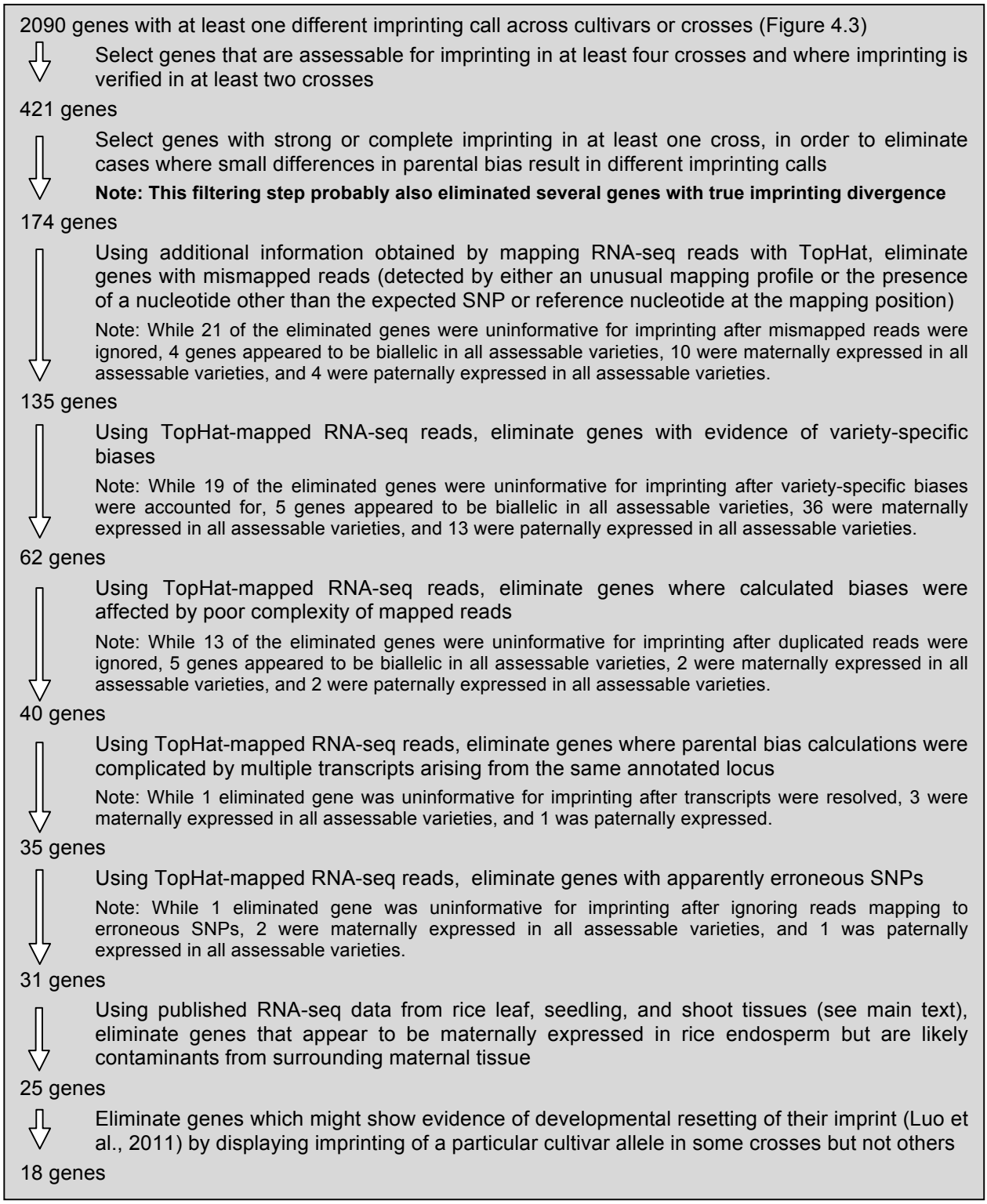


Figure 4.4 Schematic illustration of pipeline used to detect loss or gain of imprinting across rice cultivars.

Table 4.2 Genes with a loss or gain of imprinting across the four rice cultivars. Color codes for the imprinting bias in each cross are explained at the end of the table. Nip = Nipponbare; Kit = Kitaake; MEG = maternally expressed gene; PEG = paternally expressed gene; RPKM = reads per kilobase per million mapped reads.

Gene locus ID	Annotated function	Fraction of maternally-sorted reads (Total number of sorted reads reads), and expression level (RPKM) in crosses								Imprinted in cultivar?			
		Nip × IR64	IR64 × Nip	Nip × 93-11	93-11 × Nip	IR64 × 93-11	93-11 × IR64	Nip × Kit	Kit × Nip	IR64	93-11	Kit	Nip
Os01g05510	unknown	0.11 (19 reads), 0.87 RPKM	0.02 (49 reads), 0.91 RPKM	0.05 (19 reads), 1.03 RPKM	0.58 (92 reads), 1.66 RPKM	N/A (1 read), 1.13 RPKM	N/A (0 reads), 1.90 RPKM	N/A (0 reads), 0.38 RPKM	N/A (0 reads), 1.87 RPKM	PEG	No	?	PEG
Os01g69910	calmodulin-binding transcription activator	0.61 (2318 reads), 13.10 RPKM	0.09 (991 reads), 5.88 RPKM	0.44 (1645 reads), 9.55 RPKM	0.13 (1594 reads), 9.72 RPKM	N/A (0 reads), 6.87 RPKM	N/A (0 reads), 10.74 RPKM	0.13 (8 reads), 14.79 RPKM	0.25 (8 reads), 19.24 RPKM	PEG	PEG	?	No
Os01g70710	has a heavy metal-associated domain	0.68 (1077 reads), 9.57 RPKM	0.94 (644 reads), 5.82 RPKM	0.73 (2292 reads), 19.62 RPKM	0.86 (2294 reads), 19.62 RPKM	N/A (0 reads), 66.24 RPKM	N/A (0 reads), 49.05 RPKM	N/A (0 reads), 3.39 RPKM	N/A (0 reads), 4.28 RPKM	No	No	?	MEG
Os02g44530	unknown	0.86 (137 reads), 1.62 RPKM	0.63 (235 reads), 2.41 RPKM	0.94 (127 reads), 1.14 RPKM	0.65 (117 reads), 1.28 RPKM	N/A (0 reads), 3.94 RPKM	N/A (0 reads), 5.01 RPKM	N/A (0 reads), 1.88 RPKM	N/A (0 reads), 0.35 RPKM	MEG	MEG	?	No
Os02g57200	unknown	0.05 (808 reads), 13.63 RPKM	0.08 (532 reads), 7.88 RPKM	0.04 (864 reads), 11.55 RPKM	0.79 (1408 reads), 18.76 RPKM	0.15 (108 reads), 11.73 RPKM	0.61 (180 reads), 25.45 RPKM	N/A (0 reads), 10.42 RPKM	N/A (0 reads), 14.73 RPKM	PEG	No	?	PEG
Os05g47870	unknown	0.97 (86 reads), 1.44 RPKM	0.49 (55 reads), 1.67 RPKM	0.94 (64 reads), 1.06 RPKM	0.58 (24 reads), 0.67 RPKM	N/A (0 reads), 1.05 RPKM	N/A (0 reads), 1.08 RPKM	N/A (0 reads), 1.14 RPKM	N/A (0 reads), 1.67 RPKM	MEG	MEG	?	No
Os05g49240	Homeo-domain-related	0.93 (98 reads), 1.60 RPKM	0.53 (115 reads), 1.22 RPKM	0.67 (156 reads), 2.83 RPKM	0.68 (164 reads), 2.74 RPKM	0.69 (45 reads), 3.17 RPKM	0.98 (45 reads), 2.72 RPKM	N/A (0 reads), 6.63 RPKM	N/A (0 reads), 0.87 RPKM	MEG	No	?	No
Os06g13600	HEAT repeat family protein	0.00 (2972 reads), 22.21 RPKM	0.56 (3261 reads), 24.05 RPKM	N/A (0 reads), 13.90 RPKM	N/A (0 reads), 9.34 RPKM	0.62 (3596 reads), 48.27 RPKM	0.09 (1811 reads), 25.71 RPKM	N/A (0 reads), 12.27 RPKM	N/A (0 reads), 12.56 RPKM	No	PEG	?	PEG

Gene locus ID	Annotated function	Fraction of maternally-sorted reads (Total number of sorted reads reads), and expression level (RPKM) in crosses								Imprinted in cultivar?			
		Nip × IR64	IR64 × Nip	Nip × 93-11	93-11 × Nip	IR64 × 93-11	93-11 × IR64	Nip × Kit	Kit × Nip	IR64	93-11	Kit	Nip
Os08g20500	retro-transposon protein	1.00 (752 reads), 3.52 RPKM	0.60 (807 reads), 4.07 RPKM	0.99 (698 reads), 3.35 RPKM	0.62 (569 reads), 2.88 RPKM	N/A (0 reads), 3.31 RPKM	N/A (0 reads), 4.52 RPKM	N/A (0 reads), 6.15 RPKM	N/A (0 reads), 3.44 RPKM	MEG	MEG	?	No
Os09g24220	MSH-like DNA mismatch repair protein	0.54 (357 reads), 3.24 RPKM	0.09 (162 reads), 1.71 RPKM	0.68 (320 reads), 3.29 RPKM	0.27 (327 reads), 3.27 RPKM	0.00 (5 reads), 0.88 RPKM	0.00 (4 reads), 1.16 RPKM	N/A (0 reads), 5.61 RPKM	N/A (0 reads), 7.62 RPKM	PEG	PEG	?	No
Os10g05800	F-box protein, AtFBL4 homolog	0.09 (431 read), 6.73 RPKM	0.70 (366 reads), 5.07 RPKM	0.07 (338 reads), 5.43 RPKM	0.76 (394 reads), 7.86 RPKM	N/A (0 reads), 11.63 RPKM	N/A (0 reads), 10.81 RPKM	N/A (0 reads), 5.57 RPKM	N/A (0 reads), 5.45 RPKM	No	No	?	PEG
Os11g06650	transcription elongation factor protein	0.90 (1559 reads), 23.57 RPKM	0.63 (1380 reads), 25.44 RPKM	0.93 (1221 reads), 21.44 RPKM	0.62 (1658 reads), 28.54 RPKM	N/A (0 reads), 32.48 RPKM	N/A (0 reads), 30.43 RPKM	0.91 (77 reads), 23.15 RPKM	0.5 (2 reads), 10.10 RPKM	MEG	MEG	?	No
Os11g09329	VHS and GAT domain containing protein	0.05 (3041 reads), 23.24 RPKM	0.74 (3128 reads), 24.55 RPKM	0.05 (2420 reads), 19.60 RPKM	0.61 (1646 reads), 13.49 RPKM	N/A (0 reads), 31.07 RPKM	N/A (0 reads), 30.28 RPKM	N/A (0 reads), 14.35 RPKM	N/A (0 reads), 10.03 RPKM	No	No	?	PEG
Os11g31630	unknown	0.58 (509 reads), 1.85 RPKM	1.00 (329 reads), 1.41 RPKM	0.61 (463 reads), 1.76 RPKM	1.00 (137 reads), 0.60 RPKM	0.58 (134 reads), 2.00 RPKM	0.49 (131 reads), 1.95 RPKM	1.00 (69 reads), 1.27 RPKM	1.00 (12 reads), 0.31 RPKM	No	No	MEG	MEG
Os11g38990	peptidyl-prolyl cis-trans isomerase	0.02 (975 reads), 7.25 RPKM	0.79 (1086 reads), 8.08 RPKM	0.02 (1040 reads), 6.92 RPKM	0.74 (1565 reads), 10.83 RPKM	N/A (0 reads), 10.36 RPKM	N/A (0 reads), 12.42 RPKM	N/A (0 reads), 2.24 RPKM	N/A (0 reads), 6.00 RPKM	No	No	?	PEG
Os11g45295	retro-transposon protein	0.76 (5471 reads), 25.50 RPKM	0.99 (1430 reads), 6.83 RPKM	0.77 (7688 reads), 35.25 RPKM	0.99 (1373 reads), 6.43 RPKM	N/A (0 reads), 14.00 RPKM	N/A (0 reads), 6.65 RPKM	1.00 (3729 reads), 25.16 RPKM	1.00 (4295 reads), 23.41 RPKM	No	No	MEG	MEG

Gene locus ID	Annotated function	Fraction of maternally-sorted reads (Total number of sorted reads reads), and expression level (RPKM) in crosses								Imprinted in cultivar?			
		Nip × IR64	IR64 × Nip	Nip × 93-11	93-11 × Nip	IR64 × 93-11	93-11 × IR64	Nip × Kit	Kit × Nip	IR64	93-11	Kit	Nip
Os12g35590	unknown	1.00 (2207 reads), 14.74 RPKM	0.19 (1028 reads), 6.66 RPKM	1.00 (3002 reads), 16.02 RPKM	0.45 (751 reads), 4.10 RPKM	1.00 (30 reads), 0.92 RPKM	0.90 (59 reads), 2.28 RPKM	N/A (0 reads), 23.58 RPKM	N/A (0 reads), 13.79 RPKM	MEG	MEG	?	No
Os12g42600	has a ubiquitin carboxyl-terminal hydrolase domain	0.03 (670 reads), 9.82 RPKM	0.67 (754 reads), 12.84 RPKM	0.06 (535 reads), 8.57 RPKM	0.72 (902 reads), 14.59 RPKM	N/A (0 reads), 15.67 RPKM	N/A (0 reads), 17.77 RPKM	N/A (0 reads), 9.03 RPKM	N/A (0 reads), 10.66 RPKM	No	No	?	PEG

Key to color codes for imprinting biases:

Uninformative	Variety-specific bias	completely maternal	strongly maternal	moderately maternal	biallelic	moderately paternal	strongly paternal	completely paternal

The initial 174 candidate genes were the strongest candidates out of 421 genes that might display imprinting divergence. As just 10% (18 out of 174) of the strongest candidates could be manually verified to be cases of imprinting divergence, the pool of 421 candidate genes would likely yield between 18 to 44 candidates after manual filtering of technical artefacts. As another 29% of the 174 candidates appeared to be maternally expressed in all assessable cultivars and another 12.1% appeared to be paternally expressed in all assessable cultivars, it is probable that between 51 to 123 genes of the 421 genes might actually be maternally expressed in all assessable varieties and between 21 to 51 of the 421 genes might actually be paternally expressed in all assessable varieties.

Combining these numbers with the 196 genes that are paternally expressed in all assessable varieties and are supported by information in at least four crosses, and the 127 genes that are maternally expressed in all assessable varieties and are supported by information in at least four crosses, we estimate that the number of genes that have lost or gained imprinting across the four rice cultivars is 3.5 to 11% of the total number of genes that are imprinted in at least one rice cultivar.

Imprinting divergence due to a change in the dominant parental allele rather than a loss or gain of the primary imprint is rare

It can be hypothesized that divergence at an imprinted locus could result in a gene being paternally expressed in one cultivar and maternally expressed in another. In both cultivars, an imprint of some sort would still exist at the locus in question, but it would be necessary for either the primary imprint itself to have changed between cultivars or for the same primary imprint to be recognized differently based on other epigenetic differences. Such an instance of imprinting divergence due to a change in the direction of imprinted expression would not only shed light on

how mechanisms of parent-of-origin-specific gene silencing and parent-of-origin-specific gene activation are related but would also provide clues as to the kind of evolutionary selection occurring at that locus. In the event that a gene is maternally expressed in one cultivar and paternally expressed in another and neither cultivar's expression level is much greater than the others, one of the reciprocal crosses with the two cultivars would be expected to display biallelic expression of the locus while the other cross would show little or no expression.

The pipeline for detecting imprinting conservation and divergence described in Figure 4.3 would eliminate information needed to identify such a case of imprinting divergence as filtering for low coverage and variety-specific biases at genes was done such that if a gene was expressed below 0.15 RPKM in at least one cross of a set of reciprocal crosses, both crosses were considered uninformative for that gene. Thus, a different pipeline was used to assay for imprinting divergence associated with changes in the dominant parental allele rather than a loss or gain of imprinting (Figure 4.5). After screening through thousands of genes, we were only able to identify one gene that might display imprinting divergence due to a change in the dominant parental allele (Table 4.3). Further characterization of that nature of the primary imprint in each of the four cultivars is detailed in Chapter V.

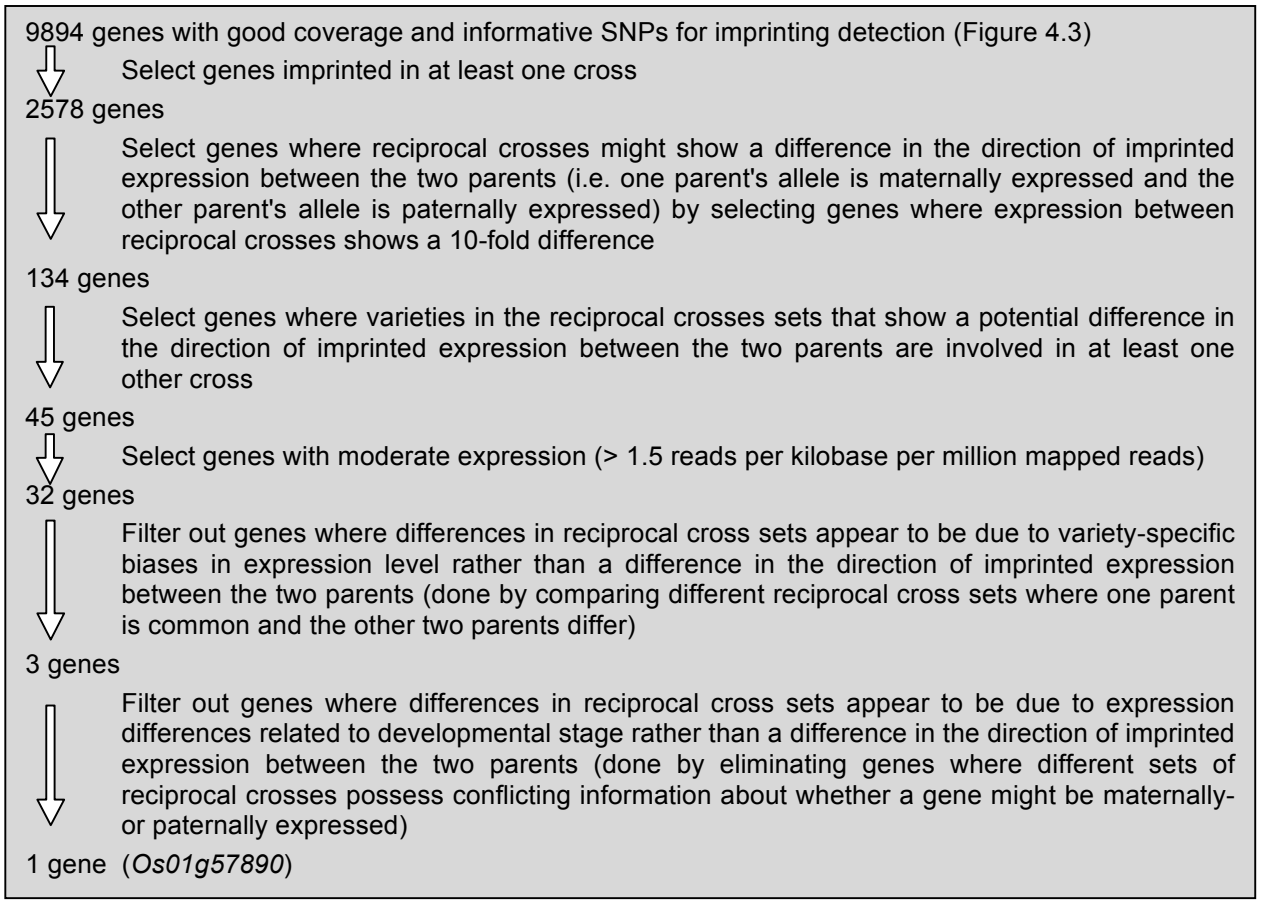


Figure 4.5 Schematic illustration of pipeline used to detect imprinting divergence that results in a change of parental bias from maternal transcript dominance to paternal transcript dominance or vice versa.

Table 4.3 *Os01g57890*, a homeobox domain containing protein, appears to be maternally expressed in the Nipponbare rice cultivar and paternally expressed in IR64 and 93-11 rice cultivars. A key to color codes for the imprinting bias in each cross is found at the end of the table. Nip = Nipponbare; Kit = Kitaake; MEG = maternally expressed gene; PEG = paternally expressed gene.

Fraction of maternally-sorted reads (Total number of sorted reads reads), and expression level (reads per kilobase per million mapped reads, RPKM) in crosses								Imprinted in cultivar?			
Nip × IR64	**IR64 × Nip**	**Nip × 93-11**	**93-11 × Nip**	**IR64 × 93-11**	**93-11 × IR64**	**Nip × Kit**	**Kit × Nip**	**IR64**	**93-11**	**Kit**	**Nip**
0.34 (30414 reads), 181.91 RPKM	0.64 (475 reads), 2.78 RPKM	0.26 (11093 reads), 73.75 RPKM	0.73 (219 reads), 1.16 RPKM	0.30 (315 reads), 13.88 RPKM	0.28 (303 reads), 11.56 RPKM	N/A (0 reads), 144.95 RPKM	N/A (0 reads), 28.96 RPKM	PEG	PEG	?	MEG

Key to color codes for imprinting biases:

Uninformative	Variety-specific bias	completely maternal	strongly maternal	moderately maternal	biallelic	moderately paternal	strongly paternal	completely paternal

Resetting of the imprint during the course of development

The only other genome-wide survey of rice imprinted genes published to date described the developmental resetting of four imprinted genes in certain crosses so that they were imprinted before 8 to 10 days after pollination (DAP) but biallelically expressed after that point (Luo et al., 2011). Paternally expressed *Os12g40520* was reset at 10 DAP in the 93-11 × Nipponbare cross but not in the Nipponbare × 93-11 cross, paternally expressed *Os12g32170* was reset at 10 DAP in the Nipponbare × 93-11 cross but not in the 93-11 × Nipponbare cross, maternally expressed *Os09g03500* was reset at 10 DAP in the Nipponbare × 93-11 cross but not in the 93-11 × Nipponbare cross, and maternally expressed *Os06g33690* was reset at 8 DAP in the Nipponbare × 93-11 cross but not in the 93-11 × Nipponbare cross.

As might be expected, our data showed these four genes were imprinted in 7- to 8-day old seeds of both reciprocal crosses with Nipponbare and 93-11. However, we identified seven other genes where different crosses provided conflicting information about whether the gene was imprinted (Table 4.4). For four of these genes, all assessable cultivars were imprinted in at least one cross. These four genes were maternally expressed *Os10g25670*, maternally expressed *Os07g28850*, paternally expressed *Os03g18080*, and paternally expressed *Os06g42910*. For the other three genes, some cultivars were not imprinted in any cross, but it is unclear whether that might be due to diverged imprinting, reset imprinting, or noisy data. These three genes were maternally expressed *Os01g59780* and *Os05g49750*, and paternally expressed *Os05g40810*.

It is possible the apparent resetting of these genes is due to undetected technical artefacts, such as those described in Figure 4.4. If they do represent true cases of developmental resetting, it would be expected that crosses with biallelic expression would possess different epigenetic marks compared to those that display imprinted expression. The resetting of a maize embryo-expressed imprinted gene was shown to be related to changes in DNA methylation at the locus (Jahnke & Scholten, 2009). In Chapter V, we test whether the same can be shown for these cases.

Table 4.4 Candidate imprinted genes that appear to display a developmental resetting of the imprint. A key to color codes for the imprinting bias in each cross is found at the end of the table. Nip = Nipponbare; Kit = Kitaake; MEG = maternally expressed gene; PEG = paternally expressed gene.

Gene locus ID	Annotated function	Fraction of maternally-sorted reads (Total number of sorted reads reads), and expression level (reads per kilobase per million mapped reads, RPKM) in crosses								Imprinted in cultivar?			
		Nip × IR64	IR64 × Nip	Nip × 93-11	93-11 × Nip	IR64 × 93-11	93-11 × IR64	Nip × Kit	Kit × Nip	IR64	93-11	Kit	Nip
Os01g59780	has an AP2 domain	0.61 (132 reads), 1.97 RPKM	0.77 (160 reads), 2.00 RPKM	0.90 (60 reads), 1.11 RPKM	0.96 (96 reads), 1.72 RPKM	N/A (0 reads), 4.70 RPKM	N/A (0 reads), 3.54 RPKM	0.55 (40 reads), 1.43 RPKM	0.73 (26 reads), 0.44 RPKM	No?	MEG	No?	MEG
Os03g18080	has a SacI homology domain	0.11 (224 reads), 1.16 RPKM	0.18 (168 reads), 1.02 RPKM	0.10 (231 reads), 1.05 RPKM	0.25 (173 reads), 0.81 RPKM	0.25 (119 reads), 1.04 RPKM	0.61 (173 reads), 1.74 RPKM	N/A (0 reads), 0.93 RPKM	N/A (0 reads), 3.17 RPKM	PEG	PEG	?	PEG
Os05g40810	has a BRCA1 C Terminus domain	0.32 (47 reads), 0.30 RPKM	0.04 (53 reads), 0.33 RPKM	0.49 (55 reads), 0.27 RPKM	0.43 (56 reads), 0.33 RPKM	N/A (0 reads), 0.42 RPKM	N/A (0 reads), 0.36 RPKM	0.57 (14 reads), 0.56 RPKM	0.00 (1 read), 1.02 RPKM	PEG	No?	?	PEG
Os05g49750	unknown	0.88 (103 reads), 4.29 RPKM	0.95 (43 reads), 2.75 RPKM	0.65 (71 reads), 3.92 RPKM	0.58 (48 reads), 3.60 RPKM	N/A (0 reads), 2.33 RPKM	N/A (0 reads), 3.08 RPKM	N/A (0 reads), 4.29 RPKM	N/A (0 reads), 5.80 RPKM	MEG	No?	?	MEG
Os06g42910	unknown	0.50 (2055 reads), 8.32 RPKM	0.01 (1935 reads), 9.14 RPKM	0.36 (2124 reads), 8.79 RPKM	0.03 (1952 reads), 8.75 RPKM	N/A (0 reads), 3.90 RPKM	N/A (0 reads), 4.44 RPKM	N/A (0 reads), 16.33 RPKM	N/A (0 reads), 12.33 RPKM	PEG	PEG	?	PEG
Os07g28850	retro-transposon protein	0.91 (1069 reads), 9.75 RPKM	0.81 (858 reads), 7.76 RPKM	0.90 (632 reads), 5.65 RPKM	0.90 (609 reads), 5.51 RPKM	N/A (0 reads), 12.57 RPKM	N/A (0 reads), 12.34 RPKM	0.70 (306 reads), 5.75 RPKM	0.71 (441 read), 8.48 RPKM	MEG	MEG	No?	MEG
Os10g25670	unknown	1.00 (131 reads), 1.12 RPKM	0.95 (57 reads), 0.35 RPKM	1.00 (114 reads), 0.91 RPKM	0.71 (73 reads), 0.56 RPKM	N/A (0 reads), 0.39 RPKM	N/A (0 reads), 0.10 RPKM	N/A (0 reads), 1.75 RPKM	N/A (0 reads), 1.18 RPKM	MEG	MEG	?	MEG

Key to color codes for imprinting biases:

Uninformative	Variety-specific bias	completely maternal	strongly maternal	moderately maternal	biallelic	moderately paternal	strongly paternal	completely paternal

6 to 8% of genes expressed in rice endosperm are imprinted

Thus, we identified 285 genes that are paternally-biased in at least two cultivars (217 of these are consistently paternally biased in at least three cultivars), 253 genes that are maternally-biased in at least two cultivars after filtering for possible seed coat contaminants (178 of these are consistently maternally biased in at least three cultivars), and one gene that exhibits either a maternal or paternal bias depending on the cultivar. A list of these genes may be found in Appendix A. We estimate that there are hundreds of other imprinted genes that we were unable to assess for imprinting.

Comparing the number of identified imprinted genes to the number assessable for imprinting, we estimate that imprinted genes constitute 6.2 to 8.1% of genes expressed in the endosperm. This is lower than our initial estimate of 10% (Figure 4.2C) as manual inspection after mapping of RNA-seq reads using the TopHat aligner identified incorrect calls due to mismapped reads, erroneous SNPs, weak variety-specific biases, poor complexity of RNA-sequencing reads at particular loci, and multiple transcripts arising from the same locus.

Functional GO term enrichment of imprinted genes reveals regulatory functions

Genes consistently imprinted among all of at least three assessable cultivars were functionally annotated with the agriGO toolkit (Z. Du, Zhou, Ling, Zhang, & Su, 2010), and gene ontology (GO) term enrichment in imprinted subsets was compared to the pool of assessable endosperm genes, as shown in Figure 4.6. 110 of the 178 maternally expressed genes and 183 of the 217 paternally expressed genes could be annotated with GO terms. Annotations were then compared to those of a list of 10493 GO-annotated genes with expression over 0.15 RPKM in the endosperm and presence of sorted reads. Enrichment was considered significant if the q-value obtained by Yekutieli (FDR under dependency) adjustment of the Fisher's exact test p-value was less than 0.05.

Conserved paternally expressed genes were similar to those in both maize (Waters et al., 2013) and *Arabidopsis thaliana* (Pignatta et al., 2014) in being enriched for functions in regulation, including that of chromatin and transcription (Figure 4.6A). Conserved maternally expressed genes were more similar to those in maize (Waters et al., 2013), being enriched for functions in transcriptional regulation, development, and signaling (Figure 4.6B). Conserved maternally expressed genes in rice appeared to be less similar to those in *Arabidopsis thaliana,* where only a weak enrichment for transcription factors was seen (Pignatta et al., 2014). However, the lack of GO annotations for 38% of conserved maternally expressed genes, compared with just 16% of paternally expressed genes that lacked annotations, suggests that the enrichment analysis performed for maternally expressed genes may not be as illustrative of the general roles of maternally expressed genes as that for paternally expressed genes. Several transposon-related proteins are maternally expressed and likely form a significant number of the unannotated group.

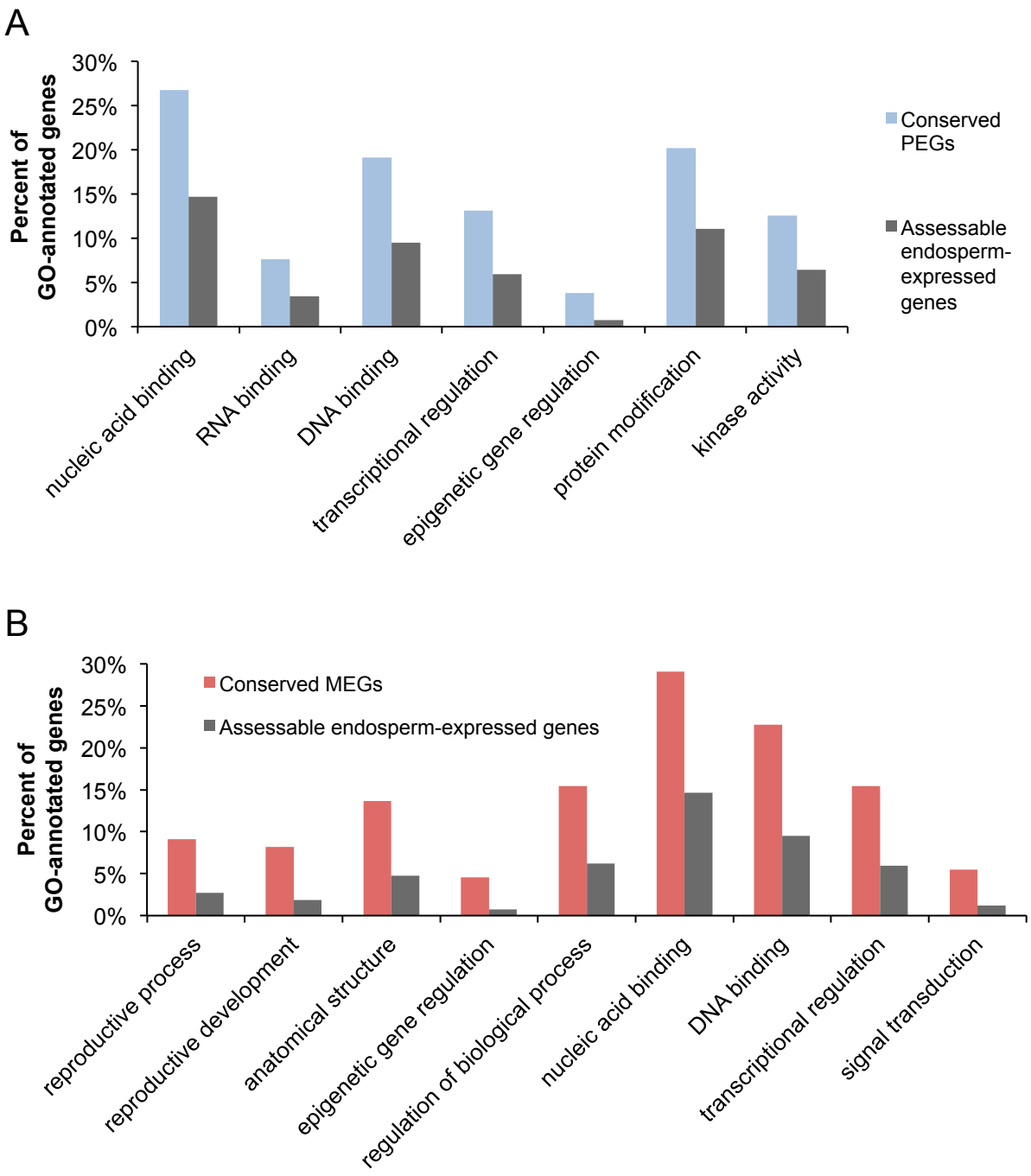


Figure 4.6 Gene ontology (GO) term enrichment of conserved imprinted genes compared to genes assessable for imprinting in rice endosperm. (A) Conserved paternally expressed genes (PEGs) were enriched for regulatory functions. (B) Conserved maternally expressed genes were enriched for roles in development, transcriptional regulation, and signal transduction.

Discussion

We find similarities between our results in rice and those from maize (Waters et al., 2013) and *Arabidopsis thaliana* (Pignatta et al., 2014) that suggest that some of the processes guiding imprinting evolution are similar in seeds with significantly different developmental programs (Agarwal, Kapoor, & Tyagi, 2011). Firstly, we find that over similarly short evolutionary time scales, the rate of imprinting divergence among rice cultivars is similar to that among maize cultivars and that among *Arabidopsis* strains. We also find that conserved imprinted genes in all three species are enriched for roles in transcriptional regulation.

This functional enrichment could be consistent with a variety of imprinting hypotheses, and it is very likely that several forces act to drive imprinting evolution and different genes may be affected by these forces differently. As the pathways that many of these transcription factors function in remain unknown, it is possible that paternally expressed transcription factors could enhance growth and nutrient acquisition by the seed while maternally expressed transcription factors inhibit growth and nutrient acquisition, thus supporting parental conflict as a driving force behind imprinting evolution (David Haig & Westoby, 1989).

However, parental conflict is predicted to drive imprinting to the complete silencing of one allele (Wilkins & Haig, 2003), and as many plant imprinted genes display only moderate biases, a dosage hypothesis for imprinting evolution has been proposed (Dilkes & Comai, 2004). Indeed, we observed that 48 (22%) of the 217 paternally expressed imprinted genes whose imprinting is conserved across all assessable rice varieties remained moderately biased across thousands of years of divergence. Another group has even reported conservation of moderate imprinting biases across the much longer evolutionary distance between rice and maize (Waters et al., 2013). Our observation that *Os01g57890* evolved different directions of parental bias in different cultivars also argues against the role of parental conflict in driving its evolution. The dosage hypothesis is related to the idea that key regulatory proteins themselves need to be strictly regulated. Based on lags between dosage sensing and subsequent induced dosage regulation, it can be predicted that it is easier to fine-tune the expression of a gene expressed from one or two copies than a gene expressed from three copies. Thus, mutations and epimutations that result in imprinting will be advantageous at such loci. It is likely that the need for dosage regulation affects plant imprinting more than mammalian imprinting, as mammalian imprinting occurs in various diploid tissues while plant imprinting is mostly confined to the triploid endosperm.

Our data provide less support for maternal-offspring coadaptation (Wolf & Hager, 2006) as a major driving force behind imprinting. The coadaptation theory suggests that genes related to transfer and communication between the maternal parent and developing offspring physically reliant on it are likely to be maternally expressed. While the majority of imprinted genes in *Arabidopsis* appear to be maternally-biased (T. F. Hsieh et al., 2009; McKeown et al., 2011; Pignatta et al., 2014; Wolff et al., 2011), we find that a greater number of imprinted genes in rice are paternally expressed, a similar situation to that in maize (Waters et al., 2013; M. Zhang et al., 2011). It has been proposed that the greater number of paternally expressed genes in maize may be driven by the need for dosage regulation (M. Zhang et al., 2011).

At this point, it is worthwhile commenting on how our detection of maternally expressed genes was complicated by contamination of endosperm samples with maternally-derived seed

coat RNA, which we observed as maternal expression of photosynthetic proteins in the non-photosynthetic endosperm. The contamination of endosperm with maternally-derived seed coat RNA is a problem frequently dealt with in studies of endosperm gene expression in *Arabidopsis thaliana*, where seeds are small and the endosperm forms a relatively small percent of the seed, thus making it difficult to separate from the seed coat (Ibarra et al., 2012; Pignatta et al., 2014; Wolff et al., 2011). We had initially expected not to encounter this problem in rice, as the rice endosperm forms a much larger proportion of the seed. However, seeds obtained from crosses used in our experiments were smaller than wildtype seeds. This is largely due to physical damage performed by our stringent emasculation procedure rather than hybridization effects, as seeds that were a result of self-fertilization after emasculation were similar in size to seeds produced by cross-fertilization. It is possible that this increased seed coat to endosperm ratio increased the potential for contamination of endosperm tissue with seed coat RNA. It is unclear whether the apparent prevalence of maternally-biased gene expression compared to paternally-biased gene expression in *Arabidopsis* is a result of seed coat contamination that is hard to account for.

Having identified variation in imprinting among rice cultivars and possible cases of reset imprinting, our next step is to determine the changes in the imprint responsible for changes to imprinted expression. It is possible that some of the apparent variation in imprinted expression is due to unaccounted technical artefacts. Variety-specific biases alone rendered a third of the genes otherwise assessable for imprinting uninformative in certain crosses. An understanding of the genetic and epigenetic variation associated with imprinting divergence will not only verify the reliability of observed expression differences but also provide insights into the molecular mechanisms guiding the evolution of imprinting. Variation in imprinting may explain some cases of pathway misregulation and abnormal seed development in rice hybrids (Ishikawa et al., 2011) as it has in mammalian hybrids (Wolf, Oakey, & Feil, 2014).

Materials and Methods

Isolation of endosperm and embryos from rice seeds

Rice plants were grown, reciprocal crosses were performed, and seeds from self-fertilized and cross-fertilized plants were harvested as described in Chapter II. Milky stage endosperm was pipetted out from inside the seed coat of 7- to 8-day old seeds and stored in TRIzol for preparing RNA-seq libraries. Embryos were isolated after the endosperm had been collected and were washed individually through vigorous agitation in 0.5 ml of 1x Phosphate Buffered Saline solution. Individually isolated F1 seeds were verified for heterozygosity with a PCR-based assay using microsatellite marker RM1 for crosses of Nipponbare with any other cultivar and RM72 for crosses between 93-11 and IR64 cultivars (McCouch et al., 2002; http://www.gramene.org/db/markers/marker_view?marker_id=24985466).

Strand-specific RNA library construction and sequencing

Strand-specific RNA-seq libraries were constructed as described in Wang et al. (2011), beginning with ribosomal RNA-depleted total RNA rather than poly-A selected transcripts. In brief, total RNA was prepared from snap-frozen rice seedling root and shoot tissue, and snap-

frozen embryos and endosperm using TRIzol reagent (Ambion). TRIzol-extracted total RNA was then treated with DNase I (Qiagen) and purified using the RNeasy Mini kit (Qiagen) as per manufacturer's instructions. After verifying the integrity of RNA on a denaturing agarose gel, 5 μg of purified total RNA was depleted for ribosomal RNA using the Ribo-Zero rRNA removal kit for plant seed and root (Epicentre) as per manufacturer's instructions. Ribosomal-depleted total RNA was purified with the solid phase reverse immobilization (SPRI) method using AM-Pure beads (Beckman Coulter) and then fragmented using the Ambion RNA fragmentation kit (Ambion).

Fragmented products were purified by overnight precipitation in 75% ethanol and the resuspended pellet was reverse transcribed using Random Hexamer primers (Invitrogen) and the SuperScript III reverse transcriptase (Invitrogen). The resultant products of this first-strand synthesis were purified using AM-Pure beads. Second-strand synthesis proceeded with DNA polymerase I (Invitrogen) and a dUTP mix of 10 mM dA, 10 mM dC, 10 mM dG, and 20 mM dU. After purification of these second-strand synthesis products with AM-Pure beads, cDNA was end-repaired using the End Repair Mix LC (Enzymatics) while still attached to AM-Pure beads. Using XP buffer (Wang et al., 2011) for intermediate size selection and washing steps, dA-tailing with Klenow exo minus (NEB) was also performed while cDNA still remained attached to the beads. cDNA was eluted off beads for Y-shape adapter ligation with DNA quick ligase (NEB) and adapters (adapter 1: 5'A*C*ACTCTTTCCCTACACGACGCTCTTCCGAT*C*T 3'; adapter 2: 5'P-G*A*TCGGAAGAGCACACGT*C*T 3'). After purifying adapter-ligated cDNA with AM-Pure beads, dUTP excision was performed with Uracil DNA glycosylase (NEB), and products were directly amplified using Phusion Hot Start 2 DNA polymerase (NEB) and indexed pair-end primers for multiplexing.

Duplex-specific nuclease treatment of RNA-sequencing libraries

After dUTP excision and PCR, products were purified and concentrated using AM-Pure beads. Duplex-specific nuclease treatment was performed as described by Matvienko et al. (2013). In short, 200 to 300 ng of library product was denatured and hybridized in TMAC buffer (3 M TMAC, 50 mM HEPES, pH 7.5) for 5 hours at 68°C. At the end of 5 hours, duplex-specific nuclease (DSN; Evrogen) was added along with DSN buffer as instructed by the manufacturer. Resulting DSN-treated products were amplified a second time using Phusion Hot Start 2 DNA polymerase (NEB) and indexed pair-end primers for multiplexing. Final DSN-treated library products were gel-purified on a 2% agarose gel to obtain fragments between 300 to 400 bp in length for sequencing. After library quantification with a Bioanalyzer (Agilent), libraries were sequenced as 100 bp single-end reads on the Illumina HiSeq 2000 platform run by the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley.

Allele-specific mapping of reads

RNA-seq reads were mapped to cDNA models for a single chosen isoform for every gene in the MSU version 7.0 genome annotation of the Nipponbare japonica rice cultivar (Ouyang et al., 2007; International Rice Genome Sequencing Project, 2005). Mapping was performed with the Bowtie read aligner (Langmead et al., 2009) and all analyses were performed with either the Nipponbare rice reference genome (MSU 7.0) or pseudo-genomes built using the Nipponbare

reference and the list of SNPs identified in Kitaake, IR64, or 93-11. SNP lists are described in Chapter II and Chapter IV, and pseudo-genomes were built using custom perl scripts as previously described (Hsieh et al., 2011). After aligning the RNA of crosses to both parental genomes, custom Perl scripts (Hsieh et al., 2011) were used to sort reads to one or the other parental genome. Reads were assigned to a parental genome in one of two instances: (1) when a read aligned to both genomes in the same position but with a different number of mismatches, or (2) when a read aligned to one genome but not to the other. For each gene model, the total number of mapped reads and the total number of sorted reads was noted. These numbers were used to calculate parental biases and assess imprinting using custom perl and python scripts based on the algorithms described in the main text of Chapter IV.

Filtering of potential maternally-derived transcripts from the seed coat

The list of candidate maternally expressed genes was filtered for possible contamination of endosperm RNA with highly expressed transcripts from the maternally-derived seed coat. Publically available RNA sequencing reads from rice leaves (Chodavarapu et al., 2012), rice shoots and roots at various developmental stages (Secco et al., 2013), and rice seedlings (T. Lu et al., 2012) were used as a proxy for the seed coat transcriptome. Raw sequencing reads from the publically available data sets were aligned to cDNA scaffolds of rice genes using Bowtie (Langmead et al., 2009) as was done for the in-house endosperm RNA-sequencing libraries. For each annotated gene, one isoform was chosen to be a representative cDNA scaffold for mapping purposes, and the expression of the isoform was assessed by calculating the number of reads mapped per kilobase of isoform sequence, per million reads that mapped to the set of cDNA scaffolds. Where biological replicate RNA sequencing datasets were available for the same tissue-type, the RPKM counts of individual libraries were averaged together to produce a final estimate for gene expression level.

TopHat visualization of RNA-seq data

RNA-sequencing reads from all endosperm libraries generated (i.e. those of self-fertilized seeds and those of cross-fertilized seeds) were mapped to either the Nipponbare rice reference genome (MSU 7.0) or to pseudo-genomes built using the Nipponbare reference and the list of SNPs identified in Kitaake, IR64, or 93-11. Mapping was done using the TopHat read aligner (Trapnell, Pachter, & Salzberg, 2009), which performs mapping across splice junctions. Mapped reads were visualized on the Integrated Genome Viewer browser in order to assess alternative splicing and possible errors in mapping. Repetitive regions were visualized after annotation of the MSU 7.0 Nipponbare reference genome using RepeatMasker with the Viridiplantae Repbase database of repetitive sequenes.

Functional GO term enrichment

As described in the text, gene ontology (GO) term annotation and enrichment for imprinted genes was performed with the agriGO toolkit (Z. Du et al., 2010). Significant enrichment compared to the pool of endosperm-expressed genes assessable for imprinting was assessed using the q-value obtained by Yekutieli (FDR under dependency) adjustment of the Fisher's exact test p-value was less than 0.05.

Chapter V

Epigenetic and genetic mutations associated with imprinting variation reinforce models for imprinting regulation

The following chapter shall soon be submitted for publication as part of a peer reviewed article.

Contributions:
Pamela Ronald, Robert Fischer, and Daniel Zilberman conceived and designed the experiment, obtained NSF funding, and provided discussion and advice at various stages. Randy Ruan peformed rice crosses and assisted in dissecting and harvesting rice endosperm and embryo. Toshiro Nishimura wrote and modified Perl Scripts for data. Manoj and Rita Sharma, in conjunction with staff at the Joint Genome Institute, performed the resequencing of the Kitaake and IR64 rice cultivars and identification of single nucleotide polymorphisms between the cultivars and the Nipponbare genome reference. I contributed to experimental design, dissected and harvested rice endosperm and embryo, and performed all molecular biology experiments and data analysis.

Introduction

Studies in *Arabidopsis thaliana*, maize, and rice indicate that DNA methylation and histone 3 lysine 27 trimethylation (H3K27me3) constitute a large part of the epigenetic imprints that distinguish maternal and paternal alleles of imprinted plant genes (Ibarra et al., 2012; Rodrigues et al., 2013; M. Zhang et al., 2014). Perturbations in pathways regulating these two marks result in a loss of imprinting at several loci (M. Du et al., 2014; Hsieh et al., 2011; Wolff et al., 2011), suggesting that they are both necessary for proper imprinted expression.

All identified cases of parent-of-origin-specific DNA methylation in the endosperm are due to site-specific hypomethylation of maternally-inherited DNA and not paternally-inherited DNA (T. F. Hsieh et al., 2009; Rodrigues et al., 2013; M. Zhang et al., 2014). In *Arabidopsis*, this relative hypermethylation of the paternal allele and relative hypomethylation of the maternal allele is due to maintenance of DNA methylation by DNA METHYLTRANSFERASE 1 (MET1) in the lineage leading to male sperm cells and active demethylation of DNA by the DEMETER (DME) DNA glycosylase in the female central cell (Hsieh et al., 2011; Morales-Ruiz et al., 2006; Penterman et al., 2007; Xiao et al., 2003). As patterns of endosperm-specific maternal hypomethylation in rice (Rodrigues et al., 2013; Zemach, Kim, et al., 2010) and maize (Gutierrez-Marcos et al., 2006; Waters et al., 2011; M. Zhang et al., 2014) are similar to that in *Arabidopsis*, it is likely that the central cell of grasses like maize and rice experience the activity of a DME-family glycosylase. The rice *DME* homolog *ROS1A* (*Os01g11900*) has an expression pattern and mutant phenotype in reproductive tissues similar to *A. thaliana DME* (A. Ono et al., 2012), and is a good candidate for a functional rice DME analog. As DME family proteins act in the female central cell and set up maternal-paternal differences before fertilization, the maternal DNA hypomethylation they effect is referred to as a primary imprint (Barlow, 1994).

Differential histone 3 lysine 27 trimethylation (H3K27me3) marks on the other hand are established and maintained post-fertilization by the selective activity of the polycomb group complex on one parental allele and not the other (Kohler et al., 2005). Genome-wide correlation studies (Weinhofer et al., 2010) as well as investigations at individual loci (M. Du et al., 2014; Villar et al., 2009) suggest that allele-specific DNA hypomethylation is required for allele-specific H3K27me3 deposition at particular loci. Thus, it appears that the primary imprint for many polycomb-regulated paternally expressed genes is once again maternal DNA hypomethylation set up by processes in the central cell. However, some imprinted genes are not associated with endosperm-specific maternal DNA hypomethylation, suggesting either that the initial imprint is due to a yet-to-be-discovered mechanism or that epigenetic states of parental alleles in the endosperm are poor representatives of the primary imprint present in gametes (Raissig, Baroux, & Grossniklaus, 2011).

We previously showed that, unlike in *Arabidopsis thaliana* (J. Lu et al., 2012), a subset of imprinted loci in rice are associated with relatively high small RNA production in the endosperm compared to embryo and seedling tissues, and that for three assessable maternally expressed genes and one paternally expressed gene, this small RNA production occurred in a parent-of-origin-specific manner from the silenced parental allele (Rodrigues et al., 2013). We found that 24-nt expression in general is unusual in the endosperm, with most small RNAs coming from relatively few loci that were highly expressed (>25 reads per kilobase per million mapped reads); we termed these loci 'siren loci' (small-interfering RNA in endosperm). It is unclear whether such imprinted siren loci play a direct role in directing imprinted gene expression from genes they overlap, or if, instead, imprinted siren locus expression is a by-product of parent-of-origin-specific silencing.

Divergence in imprinted expression among closely-related plant lineages provides an opportunity for the identification of genetic and epigenetic features required for imprinted expression, and thus has much to contribute towards our understanding of how imprints are established and maintained. In doing so, it also elucidates the lineage-specific molecular mechanisms guiding recent imprinting evolution and offers insights into variation in imprinted expression between species. Here, we explore differences in DNA methylation and small RNA production at loci previously found to show variation in imprinting among crosses or cultivars of rice (Chapter IV). We find that differences in DNA methylation at key regulatory regions correlate with imprinting divergence, and that these differences in DNA methylation associate with retrotransposition events in some cases but appear to be pure epialleles in other cases.

Results

DNA methylation and 24-nt small RNA profiles are highly similar among cultivars

To add to the DNA methylation and small RNA data obtained from Nipponbare, Kitaake, and their reciprocal crosses (Chapters II and III), we performed bisulfite sequencing of DNA (Table 5.1) and strand-specific sequencing of small RNA (Table 5.2) obtained from dissected endosperm and embryos of 7- to 8-day old seeds of IR64, and 93-11 rice cultivars. We also performed bisulfite sequencing of DNA (Table 5.1) and strand-specific sequencing of small RNA (Table 5.3) obtained from dissected endosperm and embryos of 7- to 8-day old F1 seeds produced in the reciprocal crosses used to identify imprinted expression of the cultivars in

Chapter IV. All mapping was done either to the MSU version 7.0 genome sequence of the Nipponbare japonica rice cultivar (Ouyang et al., 2007; International Rice Genome Sequencing Project, 2005) or to pseudo-genomes for each of the other three cultivars built using the Nipponbare reference and cultivar-specific single nucleotide polymorphisms. All libraries showed a good rate of bisulfite conversion, except for Kitaake endosperm, and we were able to obtain 10 to 38 fold coverage of the genome. We observed that epigenetic features are largely conserved among the four rice cultivars (Figure 5.1), with polymorphisms occurring infrequently. The conservation of general DNA methylation among rice cultivars is indicated by highly similar profiles of DNA methylation in the embryo. As regions differentially methylated between embryo and endosperm correspond to regions of endosperm-specific maternal DNA hypomethylation (Rodrigues et al., 2013), the conservation of endosperm-specific maternal DNA hypomethylation among cultivars is indicated by similar profiles of endosperm DNA methylation at places where embryo and endosperm differ in DNA methylation.

Table 5.1 Coverage and mean DNA methylation in CG, CHG, and CHH contexts for libraries that were bisulfite-sequenced. Chloroplast CHH methylation measures cytosine non-conversion and other errors. M/P = maternal/paternal; the expected ratio is 1 for embryo and 2 for endosperm. Note that some data has been previously published by our group in *(Zemach, Kim, et al., 2010), and **(Rodrigues et al., 2013).

Library	Mean coverage	Nuclear CG	Nuclear CHG	Nuclear CHH	Chloroplast CHH	M/P ratio
Nipponbare endosperm*	15	36.0%	9.7%	0.65%	0.09%	N/A
Kitaake endosperm	11	48.2%	16.2%	1.45%	0.72%	N/A
IR64 endosperm	31	38.2%	8.8%	0.67%	0.20%	N/A
93-11 endosperm	34	35.9%	9.2%	0.60%	0.20%	N/A
Nipponbare x Kitaake endosperm**	15	43.6%	13.4%	0.80%	0.20%	2.06
Kitaake x Nipponbare endosperm**	14	43.9%	11.9%	0.80%	0.20%	1.96
Nipponbare x IR64 endosperm	16	42.8%	11.1%	0.68%	0.16%	2.44
IR64 x Nipponbare endosperm	33	42.8%	10.2%	0.70%	0.17%	1.76
Nipponbare x 93-11 endosperm	38	42.9%	10.8%	0.64%	0.19%	2.48
93-11 x Nipponbare endosperm	34	43.9%	10.4%	0.80%	0.18%	1.73
IR64 x 93-11 endosperm	35	40.3%	10.5%	0.58%	0.13%	2.22
93-11 x IR64 endosperm	37	40.8%	9.9%	0.58%	0.14%	1.91
Nipponbare embryo*	14	38.5%	19.9%	3.40%	0.11%	N/A
Kitaake embryo	10	61.0%	31.4%	6.28%	0.14%	N/A
IR64 embryo	31	38.0%	15.5%	2.07%	0.19%	N/A
93-11 embryo	32	38.2%	16.8%	2.12%	0.18%	N/A
Nipponbare x Kitaake embryo**	14	47.0%	23.0%	2.50%	0.20%	1.03
Kitaake x Nipponbare embryo**	11	46.8%	22.4%	2.80%	0.20%	0.97
Nipponbare x IR64 embryo	32	47.7%	21.0%	3.00%	0.17%	1.18
IR64 x Nipponbare embryo	31	46.2%	20.7%	3.56%	0.17%	0.85
Nipponbare x 93-11 embryo	33	47.5%	20.8%	3.21%	0.18%	1.21
93-11 x Nipponbare embryo	33	47.9%	21.1%	3.34%	0.19%	0.84
IR64 x 93-11 embryo	36	45.86	19.72	3.16	0.14	1.07
93-11 x IR64 embryo	37	46.41	19.78	3.11	0.15	0.93

Table 5.2 Summary table of small RNA sequencing of embryo and endosperm tissues from IR64 and 93-11 indica cultivars. Data for Nipponbare and Kitaake previously published by our group in (Rodrigues et al., 2013) are found in table 3.1 of Chapter III.

Small RNA library	Total number of sequenced reads	Size class	Number of reads in size class	Percent size class forms of total reads	Reads that aligned	Percent of reads that aligned	Verification of sorting against the Nipponbare genome		Verification of sorting against the other indica genome (i.e. 93-11 for IR64, and IR64 for 93-11)	
							Number of reads that sorted correctly	Percent aligned reads that sorted correctly	Number of reads that sorted correctly	Percent aligned reads that sorted correctly
IR64 embryo	182416980	19 nt	3354909	1.84%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	5381477	2.95%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	9361561	5.13%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	7977134	4.37%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	30815931	16.89%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	119268045	65.38%	99931549	83.79%	14686833	97.96%	3890924	93%
		25 nt	3298730	1.81%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	2959193	1.62%	N/A	N/A	N/A	N/A	N/A	N/A
IR64 endosperm	111769073	19 nt	1684302	1.51%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	5551725	4.97%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	20567470	18.40%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	10776172	9.64%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	16733538	14.97%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	52146214	46.66%	44094426	84.56%	5745788	98.07%	1507578	92%
		25 nt	1656474	1.48%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	2653178	2.37%	N/A	N/A	N/A	N/A	N/A	N/A
93-11 embryo	143485768	19 nt	2237713	1.56%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	4081945	2.84%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	7319834	5.10%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	7881734	5.49%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	29173054	20.33%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	86558133	60.33%	72467754	83.72%	10590135	97.99%	2812621	92%
		25 nt	3004657	2.09%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	3228698	2.25%	N/A	N/A	N/A	N/A	N/A	N/A
93-11 endosperm	139775767	19 nt	2536430	1.81%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	6423779	4.60%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	16169143	11.57%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	7548241	5.40%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	21014734	15.03%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	79857549	57.13%	24587038	30.79%	2809885	97.90%	783107	95%
		25 nt	2581782	1.85%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	3644109	2.61%	N/A	N/A	N/A	N/A	N/A	N/A

Table 5.3 Summary table of small RNA sequencing of endosperm tissues from crosses involving indica (IR64 and 93-11) and japonica (Nipponbare) rice cultivars. Data for Nipponbare and Kitaake reciprocal crosses previously published by our group in (Rodrigues et al., 2013) are found in table 3.1 of Chapter III.

Cross used to generate small RNA library	Total number of sequenced reads	Size class	Number of reads in size class	Percent size class forms of total reads	Reads that aligned	Percent of reads that aligned	Number of reads that sorted to maternal parent	Percent aligned reads that sorted to maternal parent	Number of reads that sorted to paternal parent	Percent aligned reads that sorted to paternal parent
Nipponbare x IR64	134371087	19 nt	3403016	2.53%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	7834869	5.83%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	17920143	13.34%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	10239303	7.62%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	18204146	13.55%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	61251551	45.58%	57908902	94.54%	3732108	6.44%	2600635	4.49%
		25 nt	4885392	3.64%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	10632667	7.91%	N/A	N/A	N/A	N/A	N/A	N/A
IR64 x Nipponbare	124705868	19 nt	2604955	2.09%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	6472313	5.19%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	13996683	11.22%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	7728305	6.20%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	14443473	11.58%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	66284386	53.15%	27398857	41.34%	1636270	5.97%	931008	3.40%
		25 nt	4283138	3.43%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	8892615	7.13%	N/A	N/A	N/A	N/A	N/A	N/A
Nipponbare x 93-11	120332621	19 nt	2360907	1.96%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	5722978	4.76%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	16533285	13.74%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	7574637	6.29%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	15016812	12.48%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	64650984	53.73%	61730695	95.48%	4060950	6.58%	2988387	4.84%
		25 nt	2871998	2.39%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	5601020	4.65%	N/A	N/A	N/A	N/A	N/A	N/A
93-11 x Nipponbare	113311620	19 nt	5248532	4.63%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	9864998	8.71%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	22299329	19.68%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	8527217	7.53%	N/A	N/A	N/A	N/A	N/A	N/A

Cross used to generate small RNA library	Total number of sequenced reads	Size class	Number of reads in size class	Percent size class forms of total reads	Reads that aligned	Percent of reads that aligned	Number of reads that sorted to maternal parent	Percent aligned reads that sorted to maternal parent	Number of reads that sorted to paternal parent	Percent aligned reads that sorted to paternal parent
93-11 x Nipponbare	113311620	23 nt	12601299	11.12%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	38393582	33.88%	32391801	84.37%	1815126	5.60%	1052771	3.25%
		25 nt	4231982	3.73%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	12144681	10.72%	N/A	N/A	N/A	N/A	N/A	N/A
IR64 x 93-11	109566054	19 nt	6009522	5.48%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	9294927	8.48%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	20282577	18.51%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	9025119	8.24%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	13278039	12.12%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	34658164	31.63%	29902602	86.28%	472865	1.58%	363701	1.22%
		25 nt	4385922	4.00%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	12631784	11.53%	N/A	N/A	N/A	N/A	N/A	N/A
93-11 x IR64	164339930	19 nt	8271612	5.03%	N/A	N/A	N/A	N/A	N/A	N/A
		20 nt	10947704	6.66%	N/A	N/A	N/A	N/A	N/A	N/A
		21 nt	23449215	14.27%	N/A	N/A	N/A	N/A	N/A	N/A
		22 nt	10881487	6.62%	N/A	N/A	N/A	N/A	N/A	N/A
		23 nt	17357370	10.56%	N/A	N/A	N/A	N/A	N/A	N/A
		24 nt	69245052	42.14%	37065394	53.53%	572071	1.54%	452587	1.22%
		25 nt	5922053	3.60%	N/A	N/A	N/A	N/A	N/A	N/A
		Other	18265437	11.11%	N/A	N/A	N/A	N/A	N/A	N/A

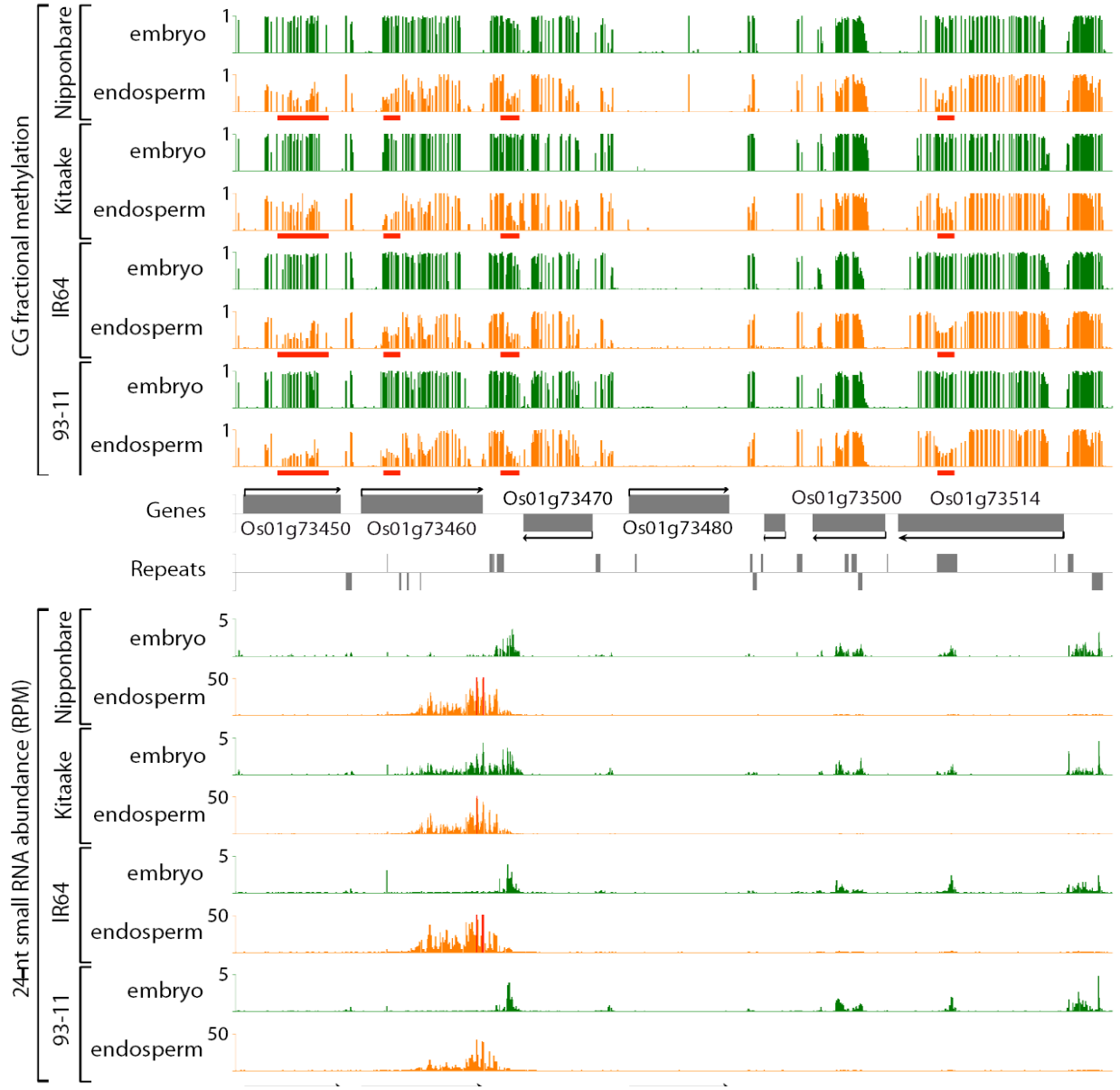


Figure 5.1. Snapshot of fractional CG methylation and 24-nt small RNA abundance in reads per million (RPM) in rice embryo and endosperm of japonica (Nipponbare and Kitaake) and indica (IR64 and 93-11) cultivars of rice. Differentially methylated regions identified between embryo and endosperm are underlined in red. Note the different scales for embryo and endosperm small RNA.

DNA methylation and/or hypomethylation polymorphisms correlate with imprinting loss or gain for 16 of 19 genes while the role of imprinted small RNAs in imprinting regulation seems minimal

We examined DNA methylation and small RNA expression patterns at the 18 genes with apparently diverged imprinting identified in Chapter IV, plus an additional gene, *Os11g16590*, that was excluded from our original analysis on account of not having a strong imprinting call. Even though *Os11g16590* was only moderately paternally-biased in all crosses that showed

imprinting, the difference in parental bias between imprinted and paternally-biased states was significant enough to make it a good candidate for further testing (difference in maternally-sorted read fractions of at least 0.4 and Fisher's exact test p-value < 0.01). It would be interesting to note if the epigenetic or genetic variation responsible for its imprinting divergence differs from that of paternally expressed genes that are strongly biased in multiple crosses, such as *Os02g57200*, *Os06g13600*, *Os10g05800*, *Os11g09329*, *Os11g38990* and *Os12g42600*.

Most of the candidate genes for imprinting divergence (18 of the 19) did not show significant endosperm 24-nt small RNA production within the gene and flanking regions up to 5 kb away. The only locus with significant 24-nt small RNA production was *Os11g09329*, which is paternally expressed in Nipponbare and biallelically expressed in IR64 and 93-11. However, small RNA production at this locus differs from that of imprinted siren loci (Rodrigues et al., 2013) in that small RNAs are produced at a lower level (10 times less than that of most siren loci) and are confined to a large intronic copia retrotransposon rather than non-repetitive intergenic regions and coding sequence. We did not have enough sorted reads to detect parental biases of the small RNAs at this locus and so we are unable to draw any conclusions about their role in mediating imprinted expression in the endosperm. Further investigation of the *Os11g09329* locus, described later in this chapter, reveals that the large intronic copia retrotransposon is absent in IR64 and 93-11 indica cultivars. It is possible that small RNA production at this retrotransposon might be related to its young age as there is evidence that it might be a recent transposition event in the Nipponbare-Kitaake lineage. Based on these data, the role of small RNAs in imprinting divergence seems indirect or minimal.

In contrast, DNA methylation or hypomethylation polymorphisms were seen at 16 of the 19 genes (Table 5.4). Imprinting at divergent maternally expressed genes was mainly associated with polymorphisms in the promoter (6 of 7 genes), transcription start site (5 of 7 genes), and downstream regulatory region (4 of 7 genes), and associated to a lesser extent with polymorphisms in the gene body (2 of 7 genes, both of which has transcriptional start site polymorphisms too) and transcription termination site (1 gene of 7 genes, which happened to have polymorphisms along the entire length of the locus). On the other hand, imprinting at the divergent paternally expressed genes was associated with polymorphisms in the promoter (6 of 9 genes), gene body (4 of 9 genes, one of which was also associated with a promoter polymorphism) and 3' downstream region (3 of 9 genes, all of which were also associated with a promoter polymorphism). These findings support our previous results (Rodrigues et al., 2013) in indicating the importance of the promoter region in imprinting of both maternally- and paternally expressed genes, and then the specific roles of the transcription start site and 3' downstream region in maternally expressed genes and that of the gene body in paternally expressed genes.

It is possible that imprinting divergence at the 3 genes without apparent variation in DNA methylation (*Os01g70710*, *Os05g47870*, and *Os11g16590*) is due to epigenetic changes associated with insertions or deletions. It is also possible that some of the 14 genes that did show epigenetic variation in DNA methylation might also possess insertions and deletions that correlate with imprinting divergence. To test this, we chose six of the 19 genes to be analyzed for large-scale genetic variation: *Os01g05510*, *Os02g57200*, *Os11g09329*, *Os11g16590*, *Os11g45295*, and *Os12g35590*. Of these six genes, *Os11g16590* did not appear to be associated with any significant epigenetic variation.

Table 5.4 Variation in DNA methylation at genes with a loss or gain of imprinting across the four rice cultivars. Nip = Nipponbare; Kit = Kitaake; TSS = Transcription start site; TTS = Transcription termination site; MEG = Maternally expressed gene; PEG = paternally expressed gene. Promoter and downstream regulatory regions were estimated to be within 5 kb of the TSS and TSS, respectively. **Os11g16590* was excluded from the original list described in Chapter IV but added here for interest.

Gene locus ID	Annotated function	Imprinted in cultivar?				Description of the DNA methylation pattern at imprinted alleles relative to the DNA methylation pattern at biallelic alleles				
		IR64	93-11	Kit	Nip	Promoter	TSS	Gene body	TTS	Downstream region
Os01g05510	unknown	PEG	No	?	PEG	Negligible variation	Negligible variation	Gain of hypo-methylation	Negligible variation	Negligible variation
Os01g69910	calmodulin-binding transcription activator	PEG	PEG	?	No	Gain of methylation that is the site of maternal hypo-methylation	Negligible variation	Negligible variation	Negligible variation	Negligible variation
Os01g70710	has a heavy metal-associated domain	No	No	?	MEG	Negligible variation	Negligible variation	Negligible variation	Negligible variation	Negligible variation
Os02g44530	unknown	MEG	MEG	?	No	Negligible variation	Negligible variation	Negligible variation	Negligible variation	Loss of methylation at a 300 bp site that is not maternally hypo-methylated
Os02g57200	unknown	PEG	No	?	PEG	Negligible variation	Negligible variation	Loss of methylation at a 500 bp site that is maternally hypo-methylated	Negligible variation	Negligible variation
Os05g47870	unknown	MEG	MEG	?	No	Negligible variation	Negligible variation	Negligible variation	Negligible variation	Negligible variation
Os05g49240	homeo-domain-related	MEG	No	?	No	Gain of methylation, but no accompanying maternal hypo-methylation	Gain of methylation, but no accompanying maternal hypo-methylation	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Negligible variation

Gene locus ID	Annotated function	Imprinted in cultivar?				Description of the DNA methylation pattern at imprinted alleles relative to the DNA methylation pattern at biallelic alleles				
		IR64	93-11	Kit	Nip	Promoter	TSS	Gene body	TTS	Downstream region
Os06g13600	HEAT repeat family protein	No	PEG	?	PEG	Gain of methylation, but no accompanying maternal hypo-methylation	Negligible variation	Negligible variation	Negligible variation	Negligible variation
Os08g20500	retro-transposon protein	MEG	MEG	?	No	Negligible variation	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Negligible variation	Gain of methylation with accompanying maternal hypo-methylation
Os09g24220	MSH-like DNA mismatch repair protein	PEG	PEG	?	No	Gain of hypo-methylation at some sites and loss of methylation at others	Negligible variation	Negligible variation	Negligible variation	Negligible variation
Os10g05800	F-box protein, AtFBL4 homolog	No	No	?	PEG	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Negligible variation	Negligible variation	Gain of methylation, but no accompanying maternal hypo-methylation
Os11g06650	transcription elongation factor protein	MEG	MEG	?	No	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Negligible variation	Negligible variation	Loss of methylation at a site associated with maternal hypo-methylation
Os11g09329	VHS and GAT domain containing protein	No	No	?	PEG	Negligible variation	Negligible variation	Some loss of methylation not associated with hypo-methylation	Negligible variation	Negligible variation
Os11g31630	unknown	No	No	MEG	MEG	Gain of methylation with accompanying maternal hypo-methylation	Gain of methylation with accompanying maternal hypo-methylation	Gain of methylation with accompanying maternal hypo-methylation	Loss of methylation not associated with maternal hypo-methylation	Loss of some methylation not associated with maternal hypo-methylation; gain of hypo-methylation at other sites
Os11g38990	peptidyl-prolyl cis-trans isomerase	No	No	?	PEG	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Gain of methylation, but no accompanying maternal hypo-methylation

Gene locus ID	Annotated function	Imprinted in cultivar?				Description of the DNA methylation pattern at imprinted alleles relative to the DNA methylation pattern at biallelic alleles				
		IR64	93-11	Kit	Nip	Promoter	TSS	Gene body	TTS	Downstream region
Os11g45295	retro-transposon protein	No	No	MEG	MEG	Gain of methylation with accompanying maternal hypo-methylation	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Negligible variation	Negligible variation
Os12g35590	unknown	MEG	MEG	?	No	Gain of methylation with accompanying maternal hypo-methylation	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Negligible variation	Negligible variation
Os12g42600	has a ubiquitin carboxyl-terminal hydrolase domain	No	No	?	PEG	Gain of methylation with accompanying maternal hypo-methylation	Negligible variation	Negligible variation	Negligible variation	Gain of methylation with accompanying maternal hypo-methylation
Os11g16590*	ATP-dependent Clp protease	PEG	No	?	PEG	Negligible variation	Negligible variation	Negligible variation	Negligible variation	Negligible variation

Cultivar-specific retrotransposon presence correlates with imprinting variation at three of six genes assessed for genetic variation

In order to assess the extent of genetic variation at six loci that showed imprinting divergence among rice cultivars, we designed primers to PCR amplify fragments of genomic DNA that spanned the gene of interest and its flanking regions up till the transcription termini of the neighboring genes. We obtained 1.7 kb overlapping fragments and performed Sanger sequencing to verify known SNPs and identify new SNPs and indel events. By building contigs from the overlapping fragments, we obtained the genetic sequence of the six loci of interest in all four varieties.

We did not detect any significant insertions or deletions in regulatory regions that correlated with imprinting status for three of the six genes. These three genes were already associated with DNA methylation and/or hypomethylation polymorphisms: *Os01g05510*, *Os11g45295*, and *Os12g35590*. We did detect some genetic variation outside of the expected regulatory regions, but it is unclear whether they may affect imprinted expression. For example, the candidate MEG *Os11g45295* displayed one small indel in the gene body that did not correlate with imprinting and another larger indel in the 3' downstream area that correlated with biallelic gene expression. The relevance of the larger indel for *Os11g45295* is unknown as it

seems to occur in the promoter of a downstream truncated gene, *Os11g45320* (Figure 5.2). The most convincing variation that might be the cause of imprinting divergence is a DNA methylation epiallele around the transcription start site of the isoform of *Os11g45295* that is expressed in endosperm. Although this occurs in close proximity to the small gene body indel, the two do not correlate with each other, suggesting that the methylation difference is a true epiallele unrelated to large indels.

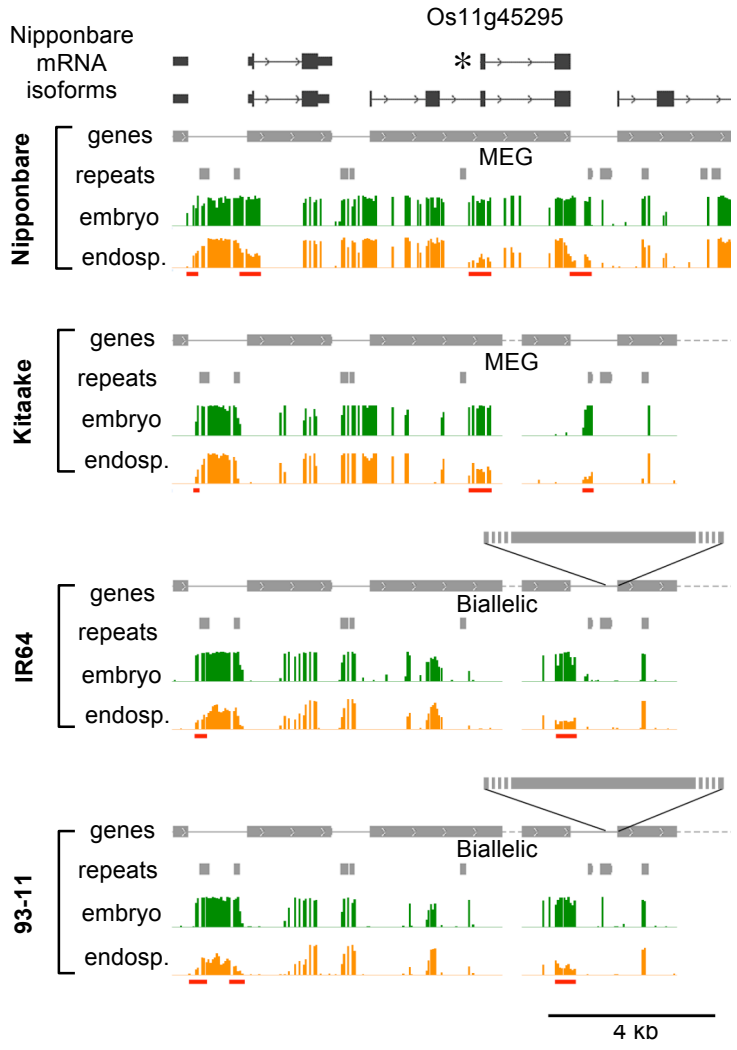


Figure 5.2 Epigenetic and genetic variation among cultivars at *Os11g45295* shows that a DNA methylation epiallele around the transcription start site of the endosperm-expressed isoform of the gene (shown with an asterisk) correlates with imprinting divergence. Fractional CG methylation is shown for embryo (green) and endosperm (orange) in each cultivar, and sites of endosperm-specific hypomethylation that correspond to maternal hypomethylation are underlined in red. Deletion events are indicated as a dotted line along the gene annotation while insertion events are indicated as wedges above the gene annotation.

The paternally expressed gene *Os01g05510* was deleted in Kitaake, but no indel events were found among Nipponbare, 93-11 and IR64 (Figure 5.3A). The paternal bias of this gene in Nipponbare and IR64 appears to be related to endosperm-specific hypomethylation in the second half of the gene body. The 93-11 allele shows a gain of additional methylation in the middle of this region, and although flanking regions are hypomethylated in the endosperm, the 93-11-specific methylated region is not. It is likely that maternal DNA hypomethylation in Nippon--bare and IR64 allows for specific recognition of the maternal and not paternal allele by the polycomb group, such that maternal-specific repression results in a paternal expression bias. The presence of methylation in 93-11 that is not specifically removed from the maternal allele most likely prevents recognition of the maternal allele by the polycomb group despite hypomethylation at flanking regions.

Finally, maternally-biased expression of *Os12g35590* (Figure 5.3B), in a manner similar to that of *Os11g45295* (Figure 5.2), correlates with cultivar-specific DNA methylation that undergoes maternal-specific hypomethlation. An indel corresponding to a small nearby gene, *Os12g35600*, correlates with the cultivar-specific methylation apparently responsible for imprinting. However, it is unclear how this indel might affect methylation at a site 3 kb away. As methylation profiles closer to the site of the indel do not seem to be affected by the presence or absence of the indel, it is more likely that the indel does not have an effect on the DNA methylation and hypomethylation polymorphism that correlates with imprinting and that the polymorphism is a pure epiallele unrelated to genetic variation.

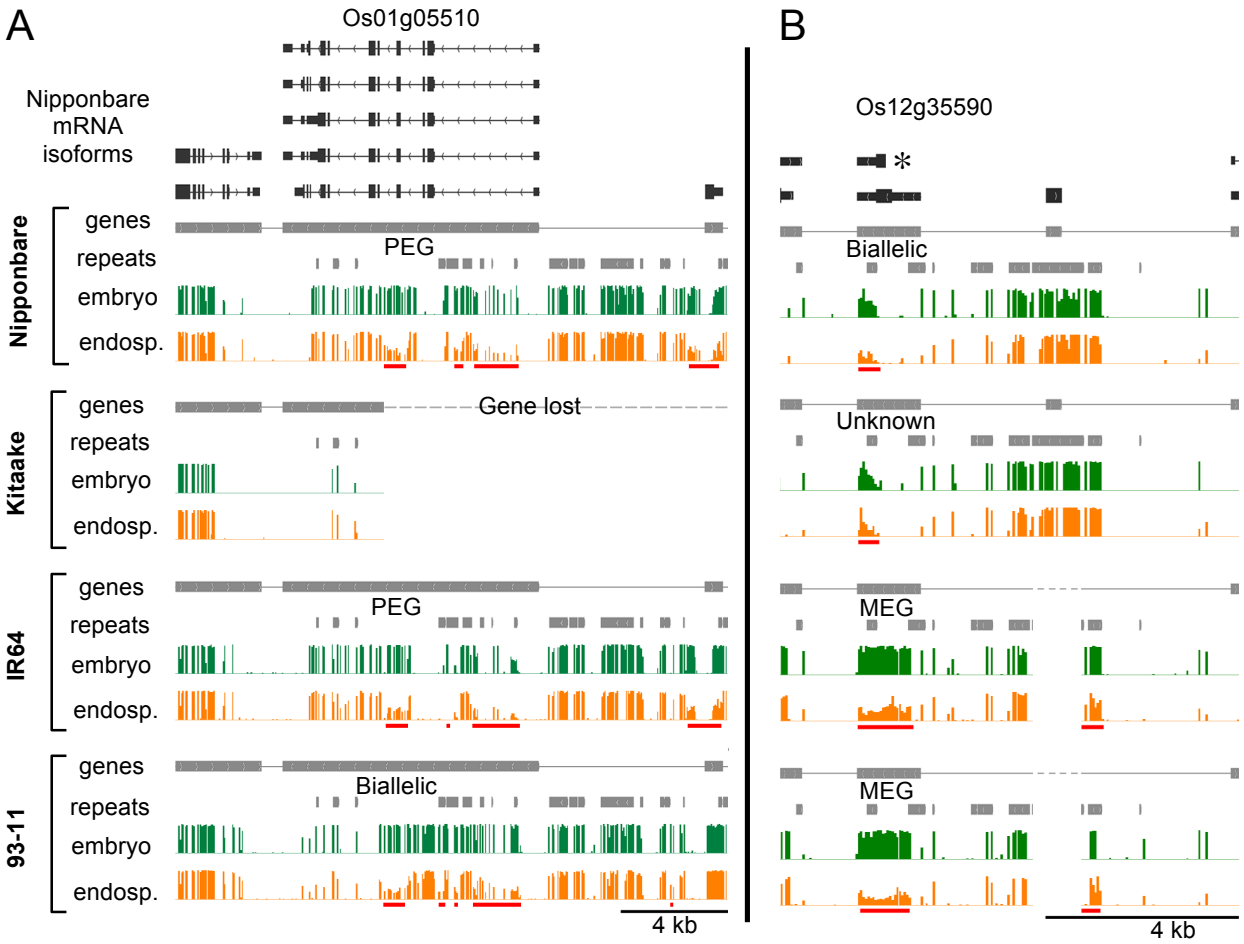


Figure 5.3 Epigenetic and genetic variation among cultivars at (A) *Os01g05510* and (B) *Os12g35590* shows that DNA methylation epialleles unrelated to large genetic variation correlate with imprinting divergence. For paternally expressed *Os1g05510* (A) this occurs in the gene body, and for maternally expressed *Os12g35590* (B) this occurs around the transcription start site of the endosperm-expressed isoform of the gene (indicated with an asterisk). Fractional CG methylation is shown for embryo (green) and endosperm (orange) in each cultivar, and sites of endosperm-specific hypomethylation that correspond to maternal hypomethylation are underlined in red. Deletion events are indicated as a dotted line along the gene annotation while insertion events are indicated as wedges above the gene annotation.

For the other three genes, we observed indels consisting of 3 kb- to 6 kb-long terminal repeat (LTR) retrotransposons (Figures 5.4A,B, 5.5). In *Os02g57200* (Figure 5.4A), a 6 kb intronic copia element was associated with the biallelic 93-11 allele and absent in paternally expressed IR64 and Nipponbare alleles. The density of SNPs between cultivars at the locus indicates that 93-11 and IR64 loci are more recently diverged from each other than the either is from Nipponbare. We could not detect solo long terminal repeats (LTRs) in IR64 and Nipponbare that might be evidence of loss of retrotransposon sequence in those lineages by genetic recombination (Tian et al., 2009). It is possible that the retrotransposon might have excised precisely, as has been described in Drosophila (Kuzin, Lyubomirskaya, Khudaibergenova, Ilyin, & Kim, 1994), however, precise retrotransposon excision has not been observed in rice (Ma & Bennetzen, 2004). Thus, it is likely that the retrotransposon inserted into the 93-11 lineage after its divergence from IR64, causing biallelic expression of a previously paternally expressed gene through disruption of the imprinting control region.

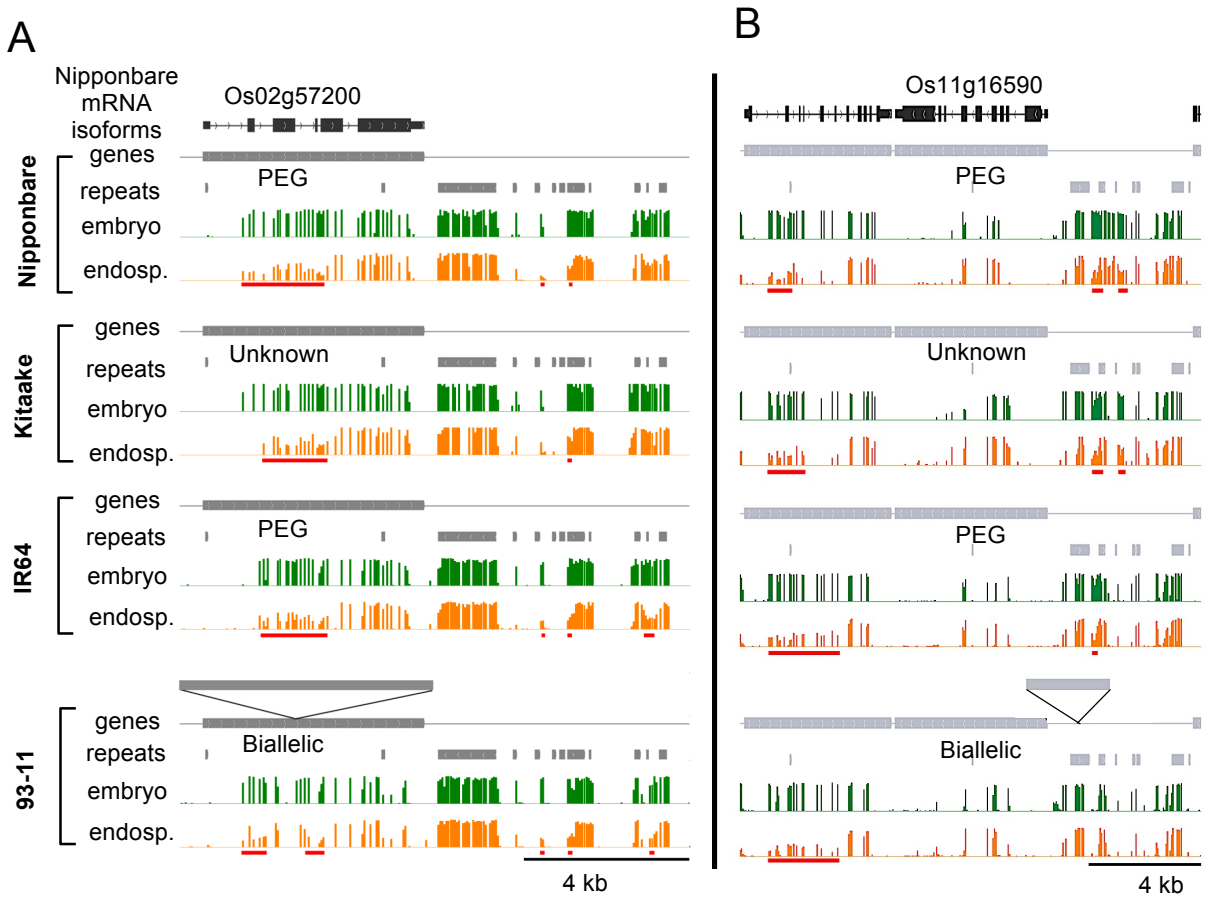


Figure 5.4 Epigenetic and genetic variation among cultivars at (A) *Os02g57200* and (B) *Os11g16590* show that 93-11-specific retrotransposon insertions correlate with imprinting divergence. In *Os02g57200* (A) the insertion occurs in the gene body, while in *Os11g16590* (B) it occurs in the promoter region. Fractional CG methylation is shown for embryo (green) and endosperm (orange) in each cultivar, and sites of endosperm-specific hypomethylation that correspond to maternal hypomethylation are underlined in red. Deletion events are indicated as a dotted line along the gene annotation while insertion events are indicated as wedges above the gene annotation.

Similarly, the 3 kb gypsy LTR retrotransposon insertion in *Os11g16590* (Figure 5.4B) is present in the promoter of the biallelically expressed 93-11 locus and is absent in the paternally expressed alleles of Nipponbare and IR64. As we could not detect solo LTRs in either Nipponbare or IR64, and SNP density indicated that 93-11 and IR64 loci were more closely related to each other than to those of the japonica cultivars, it is likely that the retrotransposon is a recent insertion in the 93-11 lineage and that its insertion resulted in a loss of imprinting. Interestingly, this paternally expressed gene does not display the endosperm-specific gene body hypomethylation that is commonly seen in association with silencing of the maternal allele and expression of the paternal allele. However, *Os11g16580*, which is downstream of it and in close proximity does display PEG-type hypomethylation in the gene body. Thus, it is possible that polycomb-mediated silencing of the maternal allele of *Os11g16580* usually silences the maternal allele of *Os11g16590* too, but the promoter insertion in the 93-11 allele is able to counteract this. Silencing of the *PHERES1* gene in *Arabidopsis thaliana* also occurs through effects of DNA methylation on polycomb activity several kb away (Villar et al., 2009). Alternatively, it is possible that *Os11g16590* is regulated solely through promoter hypomethylation rather than gene body hypomethylation. A small DMR upstream of the transposon insertion site appears to correlate with imprinting. While this small amount of methylation might not significantly contribute to the overall epigenetic profile of the locus, it is possible that it deters the binding of specific gene regulatory proteins that may be targeted to the underlying sequence.

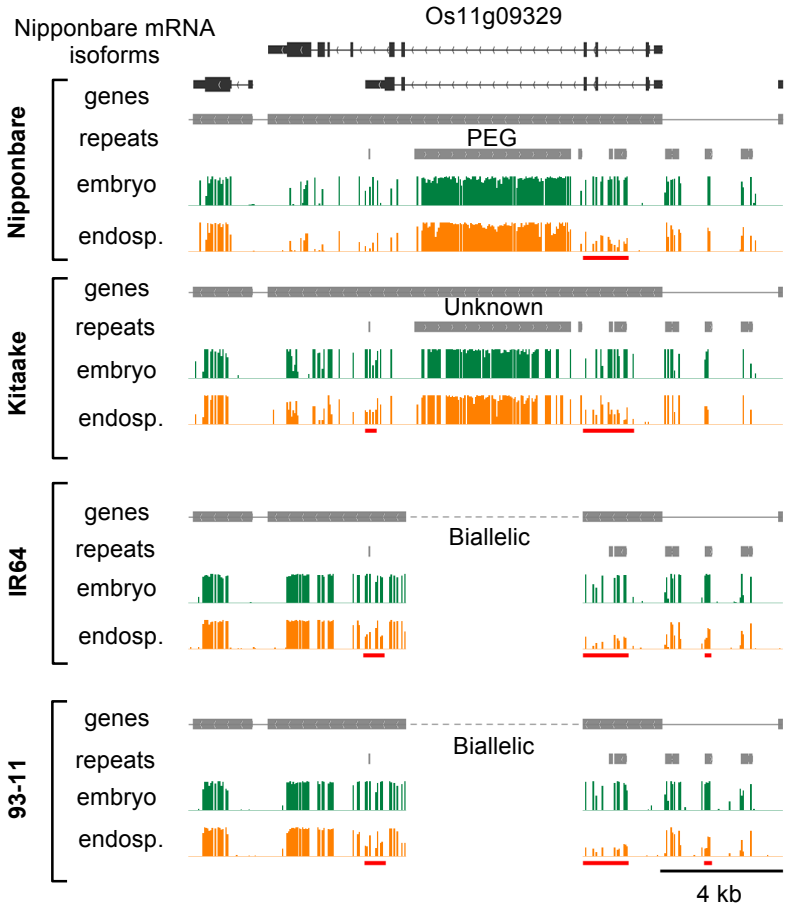


Figure 5.5 Epigenetic and genetic variation among cultivars at *Os11g09329* reveals that imprinting is associated with the presence of 5.5 kb copia retrotransposon. Note that many reads do not map uniquely to the retrotransposon sequence, so our estimation of its methylation is impaired. Fractional CG methylation is shown for embryo (green) and endosperm (orange) in each cultivar, and sites of endosperm-specific hypomethylation that correspond to maternal hypomethylation are underlined in red. Deletion events are indicated as a dotted line along the gene annotation while insertion events are indicated as wedges above the gene annotation.

Finally, in *Os11g09329* (Figure 5.5), a 5.5 kb copia intronic retrotransposon present in Nipponbare but not the indica cultivars correlates with paternal expression and imprinting of the Nipponbare allele. It is likely that the presence of the retrotransposon provides a signal for polycomb recruitment and that maternal specific hypomethylation of the retrotransposon allows silencing of the maternal allele but not paternal allele. Due to the paucity of unique reads within the intronic retrotransposon, we were unable to accurately estimate the methylation status of the gene body of *Os11g09329* (Figure 5.5) but the data suggest that there is a high likelihood that the region is methylated on both maternal and paternal alleles in the embryo and on the paternal allele in the endosperm. We could not detect solo LTRs of the retrotransposon in the indica cultivars, suggesting that the retrotransposition event did not occur in those lineages. The presence of the retrotransposon in the Kitaake cultivar indicates that the retrotransposon insertion occurred before the Nipponbare allele began to diverge from the Kitaake allele.

Comparisons of sequence identity between the two LTRs for each of the cultivar-specific copia retrotransposons in *Os11g09329* and *Os02g57200* indicate that the two retrotransposons would have inserted into their respective loci within a relatively short time of each other on an evolutionary scale. The 366 bp LTRs flanking the copia element in the 93-11 allele of *Os02g57200* possess four SNPs with respect to each other, while the 362 bp LTRs of the copia element in Nipponbare *Os11g09329* differ at two SNPs and two indels (one 18 bp indel and a second 23 bp indel). It is unclear how one insertion resulted in a loss of imprinting while the other insertion resulted in a gain of imprinting over similar time scales. A better understanding of the mechanism of polycomb group recruitment and activity is likely to shed light on this observation. A summary of the combinations of epigenetic and genetic variation that are probable causes of imprinting divergence at the six assessed loci is found in Table 5.5.

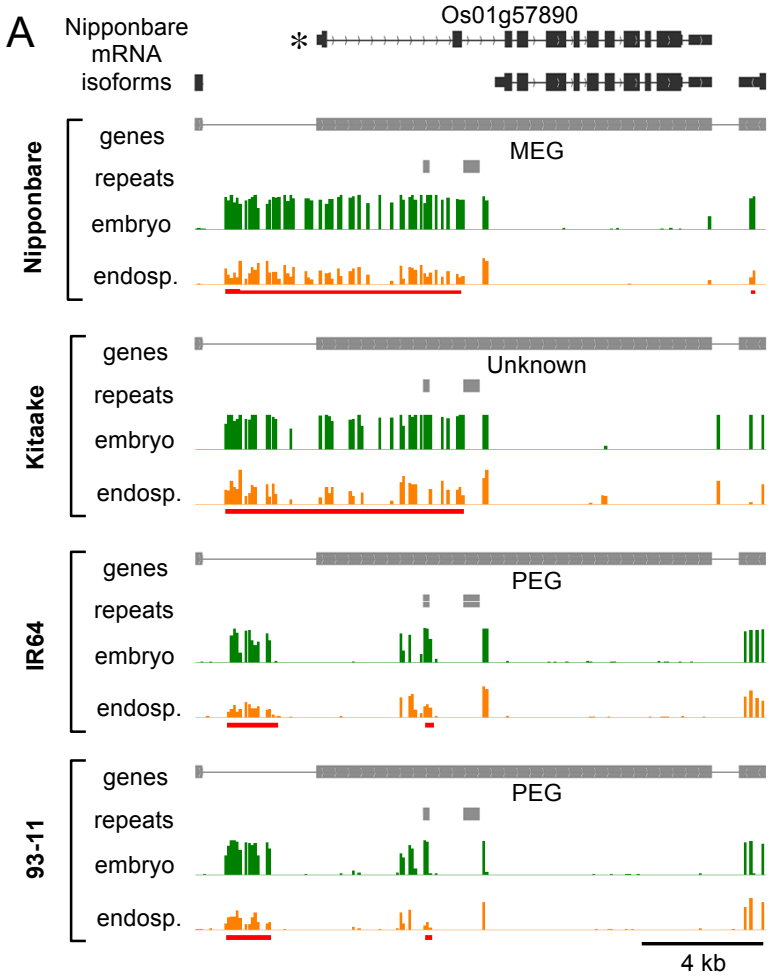
Table 5.5 Summary of epigenetic and genetic variation at imprinted alleles compared to biallelic alleles for six genes whose genetic sequence was verified by Sanger sequencing.

Gene locus	Imprinting divergence	Promoter	Transcription start site	Gene body	Transcription termination site	Down-stream
Os01g05510	Paternal and biallelic	Negligible variation	Negligible variation	**Loss of methylation**	Negligible variation	Negligible variation
Os02g57200	Paternal and biallelic	Negligible variation	Negligible variation	**6 kb intron deletion**	Negligible variation	Negligible variation
Os11g09329	Paternal and biallelic	Negligible variation	Negligible variation	**5.5 kb intron insertion**	Negligible variation	Negligible variation
Os11g16590	Paternal and biallelic	**3 kb deletion**	Negligible variation	Negligible variation	Negligible variation	Negligible variation
Os11g45295	Maternal and biallelic	**Gain of methylation**	**Gain of methylation**	Non-correlative 400 bp insert	Negligible variation	**Large deletion**
Os12g35590	Maternal and biallelic	**Gain of methylation**	**Gain of methylation**	Negligible variation	Negligible variation	Negligible variation

Epigenetic variation may explain differences in the direction of imprinted expression at *Os01g57890*

The expression pattern of the homeodomain-containing protein *Os01g57890* in reciprocal crosses seemed to indicate that it was maternally expressed in the Nipponbare rice cultivar and paternally expressed in the IR64 and 93-11 rice cultivars. In order to verify these results and identify possible causes for the difference in direction of imprinted expression, we verified the genetic sequence of cultivars using PCR amplification and Sanger sequencing, and assayed for variation in DNA methylation and small RNAs as before. We did not find any small RNA production or large indels associated with the locus, but we did find that an epiallele not caused by large genetic variation correlates with parental bias (Figure 5.6).

A



B

Promoter	Transcription start site	Gene body	Transcription termination site	Down-stream
Negligible variation	**Epiallele**	**Epiallele**	Negligible variation	Negligible variation

Figure 5.6 Epigenetic and genetic variation among cultivars at *Os01g57890* reveals that maternal bias is associated with the presence of DNA methylation and endosperm-specific hypomethylation around the transcription start site of the endosperm-expressed isoform (indicated with an asterisk). In contrast, paternal bias is associated with a lack of methylation around the transcription start site but a retention of methylation and endosperm-specific hypo-methylation at the promoter and a few sites in the gene body. (A) Snapshot of the genome in the four rice cultivars. Fractional CG methylation is shown for embryo (green) and endosperm (orange) in each of the cultivars, and sites of endosperm-specific hypomethylation due to maternal hypomethylation are underlined in red. Note that no indels were detected among the four cultivars, suggesting that imprinting and DNA methylation polymorphimsms are unrelated to large changes in genetic sequence. A summary of the assessment of epigenetic and genetic variation is provided in (B).

Based on SNP density and DNA methylation profile, it appears that the opposing biases evolved after indica and japonica subspecies diverged from each other but before the two indica cultivars, 93-11 and IR64, began to diverge. It is unknown whether each imprinted state evolved independently from an initial biallelic allele in the last common ancestor for the cultivars, or whether the allele in the last common ancestor was imprinted and one of its descendant lineages transitioned from one parental bias to the opposing bias. In either case, the finding that a gene has the capacity to rapidly evolve imprints leading to either maternal or paternal bias yields interesting insights into the mechanism of imprinting evolution.

We could not verify developmental resetting of the imprint for candidate reset genes

In their imprinting survey using reciprocal crosses between Nipponbare and 93-11 rice varieties, Luo et al. (2011) described genes that show temporal variation in imprinting between 3.5 to 10 days after pollination (DAP). They described four genes that were biallelic between 3.5 to 6 DAP and imprinted after that: maternally expressed *Os06g33640*, maternally expressed *Os09g03500*, paternally expressed *Os12g32170*, and paternally expressed *Os12g40520*. They also described four genes that were imprinted before 8 to 10 DAP and were biallelic beyond that point in at least one cross: maternally expressed *Os06g33640*, maternally expressed *Os09g03500*, paternally expressed *Os12g32170*, and paternally expressed *Os12g40520*.

As might be expected, all eight of these genes appeared to be imprinted in all assessable crosses in our data from 7 to 8 DAP seeds, even though our data included reciprocal crosses between Nipponbare and 93-11. However, we were able to identify our own list of seven imprinted genes that might exhibit temporal variation in imprinted expression (Chapter IV): maternally expressed *Os01g59780*, *Os05g49750*, *Os07g28850*, and *Os10g25670*, and paternally expressed *Os03g18080*, *Os05g40810*, and *Os06g42910*. It is possible that these genes may not truly represent cases of reset imprinting as we found that many imprinting calls in our analysis were confounded by mismapped reads, erroneous SNPs, weak variety-specific biases, poor complexity of RNA-sequencing reads at a particular locus, and contamination of endosperm samples with RNA from maternal tissues. A previous study in maize (Jahnke & Scholten, 2009) found changes in DNA methylation to be associated with a case of developmental resetting in the embryo, so we aimed to verify whether the seven genes we identified might truly represent cases of development resetting by assessing if a change in the imprint can be detected through DNA methylation and small RNA profiles.

In brief, not one of the seven genes showed significant differences in DNA methylation or small RNAs that correlated with potential developmental resetting. Six of the seven loci had no or very little 24-nt sRNA production in any of the crosses. The seventh gene, paternally expressed *Os06g42910*, was associated with expression of a siren locus in all four cultivars, though expression level varied (Figure 5.7B). 24-nt sRNA expression differed in some sets of reciprocal crosses, being greater when either of the japonica cultivars (Kitaake and Nipponbare) was used as the maternal parent. This likely reflects variety-specific differences in expression level that become more dominant if present as two maternal copies rather than one paternal copy. However, sRNA expression in the apparently reset Nipponbare x IR64 cross was similar to Nipponbare x 93-11 which displays imprinted gene expression (Figure 5.7A).

A

Gene locus ID	Annotated function	Parental bias, fraction of maternally-sorted reads (total number of sorted reads reads), and RPKM in crosses								Imprinted in cultivar?			
		Nip × IR64	IR64 × Nip	Nip × 93-11	93-11 × Nip	IR64 × 93-11	93-11 × IR64	Nip × Kit	Kit × Nip	IR64	93-11	Kit	Nip
Os06g42910	unknown	bialleic 0.50 (2055 reads), 8.32 RPKM	compl. pat 0.01 (1935 reads), 9.14 RPKM	mod. pat 0.36 (2124 reads), 8.79 RPKM	strongly pat 0.03 (1952 reads), 8.75 RPKM	N/A (0 reads), 3.90 RPKM	N/A (0 reads), 4.44 RPKM	N/A (0 reads), 16.33 RPKM	N/A (0 reads), 12.33 RPKM	PEG	PEG	?	PEG

B

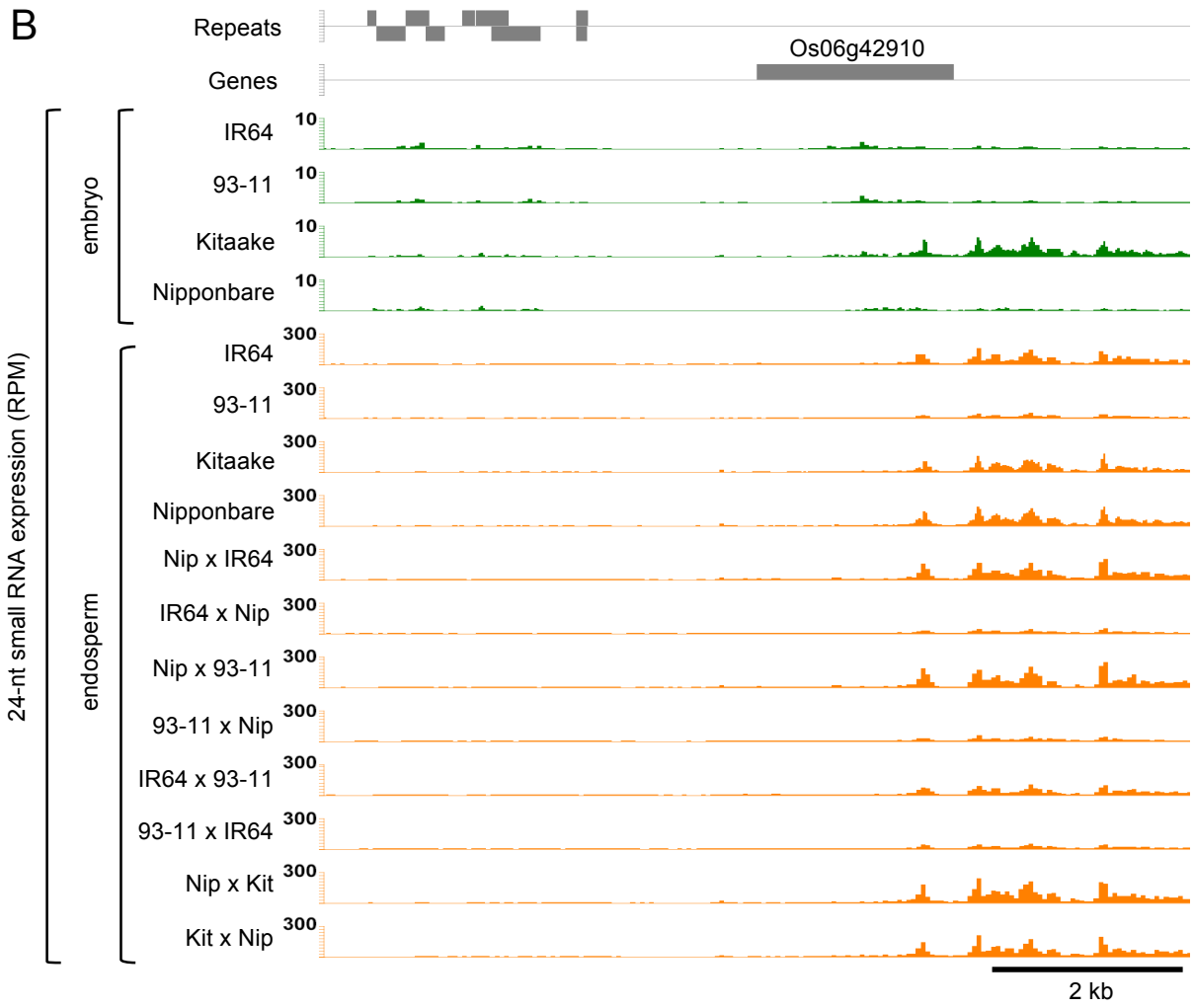


Figure 5.7 24-nt sRNA abundance does not correlate with the apparent developmental resetting of *Os06g42910*. (A) Parental biases in reciprocal crosses imply that *Os06g42910* is a conserved PEG but is biallelically expressed in the Nipponbare x IR64 cross due to developmental resetting. (B) 24-nt small RNA expression reveals that the gene overlaps a siren locus expressed in all cultivars but at different levels. Note the differing scales for endosperm and embryo. Nip = Nipponbare; Kit = Kitaake; PEG = paternally expressed gene; compl. pat = completely paternal; strongly pat = strongly paternal; mod. pat = moderately paternal; RPM = reads per million; RPKM = reads per kilobase per million mapped reads.

Analysis of the apparent resetting of *Os06g42910* reveals an unusual mechanism of paternal-specific expression

We tested whether parental biases in 24-nt small RNA expression at *Os06g42910* correlate with a potential resetting of the imprint by comparing the ratio of maternally- and paternally-sorted 24-nt small RNA reads between crosses (Figure 5.8). Once again, there were no differences between the apparently reset Nipponbare x IR64 cross and the imprinted Nipponbare x 93-11 cross, but we did notice cultivar-dependent variation that suggests that small RNAs at this locus in endosperm are more maternally-biased in IR64 and 93-11 than in Nipponbare.

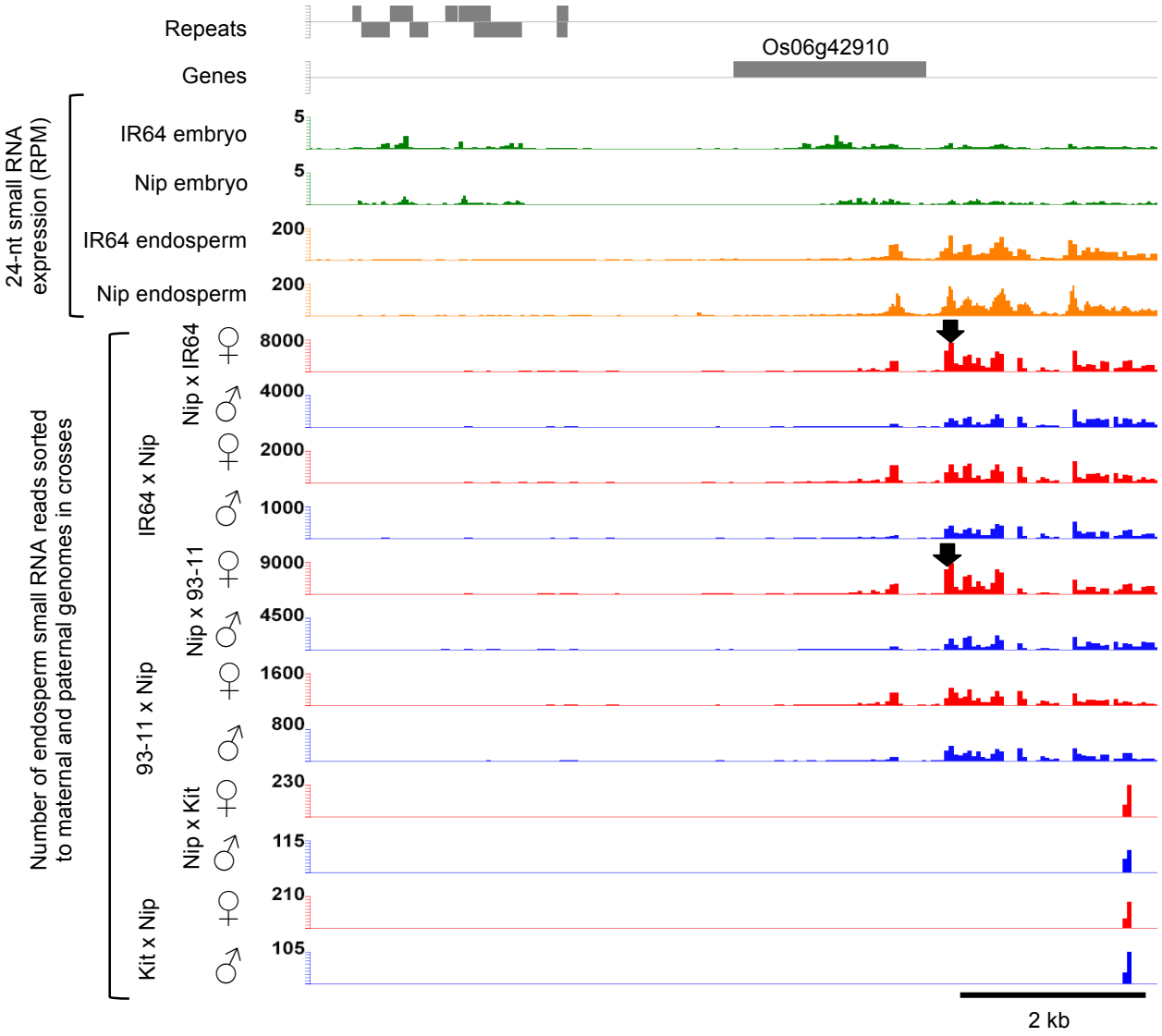


Figure 5.8 Small RNA expression at *Os06g42910* appears to be more maternally-biased in IR64 and 93-11 than in Nipponbare. Note the different scales for embryo (green) and endosperm (orange) and for maternally- (red) and paternally- (blue) sorted reads. The different scaling of maternal and paternal reads reflects the 2:1 ratio of parental genomes in the endosperm. Arrows indicate crosses that display the maternal bias suggesting that small RNAs are more maternally biased in IR64 and 93-11. Small RNAs at the same region appear to be biallelic in Nipponbare (Nip) and Kitaake (Kit) based on the other crosses.

As we previously found that genes overlapping imprinted siren loci were expressed from the opposite allele (Rodrigues et al., 2013), it is likely that this result points to a paternal bias for 93-11 and IR64 alleles and biallelic expression of the Nipponbare allele. A combination of such imprinting divergence and variety-specific biases may explain the apparent developmental resetting. We observed DNA methylation profiles at the *Os06g42910* locus in order to assess whether they corroborate variety-specific imprinting (Figure 5.9). However, DNA methylation profiles neither supported the paternal expression of indica cultivar alleles and biallelic expression in Nipponbare, nor developmental resetting in the Nipponbare x IR64 cross. It is possible that this may be due to developmental resetting in a mixed population of cells, where DNA methylation profiles represent each cell in the population equally but RNA profiles do not. It may also be that *Os06g42910* imprinting is not regulated by DNA methylation and hypomethylation. At this stage in seed development, it is unlikely that apparent paternal-expression bias is a result of contamination with paternally-derived tissue or long-lived RNAs from the sperm cell, so an imprint must distinguish parent alleles in order to cause the bias.

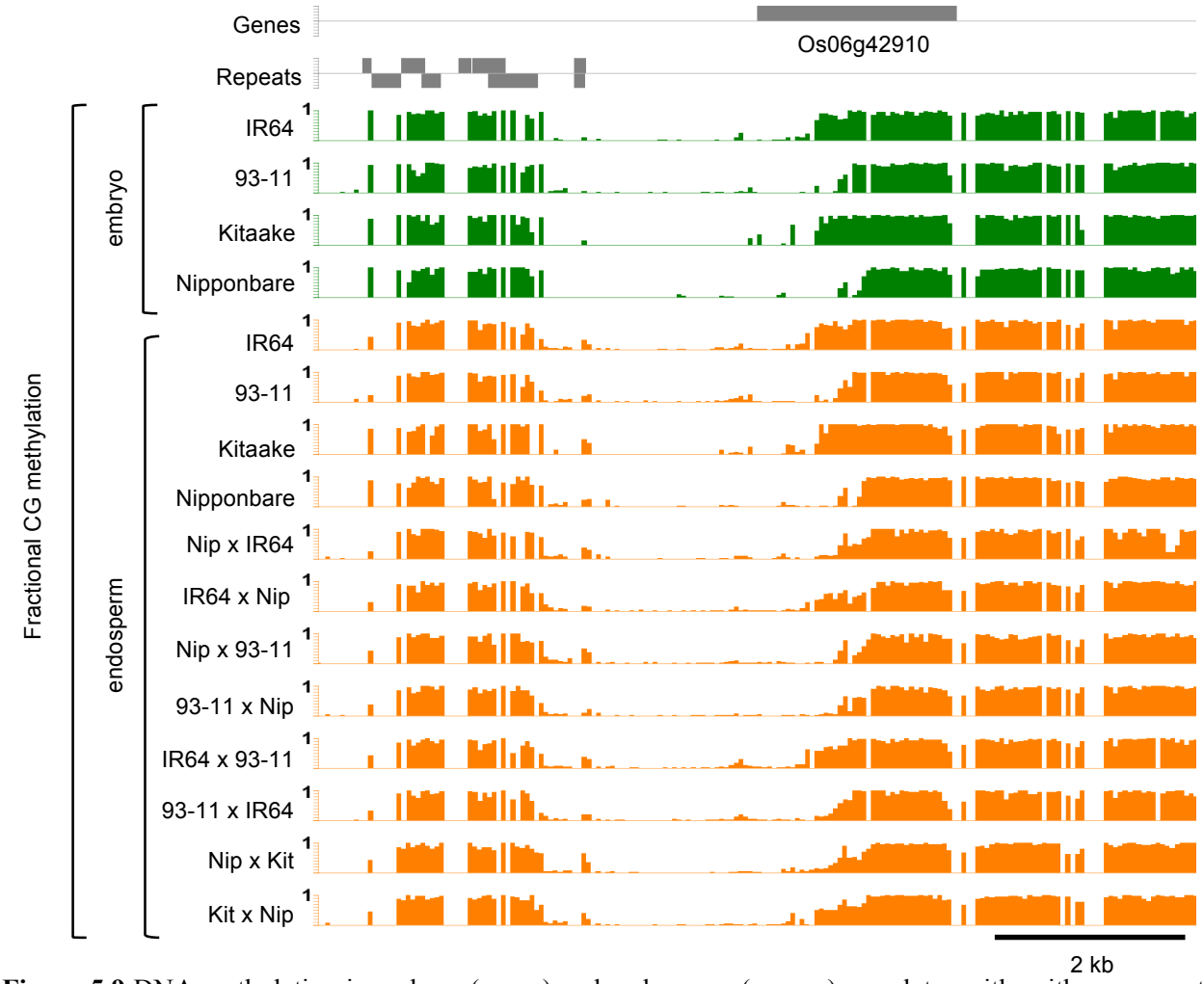


Figure 5.9 DNA methylation in embryo (green) and endosperm (orange) correlates with neither apparent developmental resetting nor potential imprinting divergence of *Os06g42910*. Nip = Nipponbare; Kit = Kitaake.

A

Gene locus ID	Annotated function	Parental bias, fraction of maternally-sorted reads (total number of sorted reads reads), and RPKM in crosses								Imprinted in cultivar?			
		Nip × IR64	IR64 × Nip	Nip × 93-11	93-11 × Nip	IR64 × 93-11	93-11 × IR64	Nip × Kit	Kit × Nip	IR64	93-11	Kit	Nip
Os01g59780	has an AP2 domain	biallelic 0.61 (132 reads), 1.97 RPKM	biallelic 0.77 (160 reads), 2.00 RPKM	strongly mat 0.90 (60 reads), 1.11 RPKM	strongly mat 0.96 (96 reads), 1.72 RPKM	N/A (0 reads), 4.70 RPKM	N/A (0 reads), 3.54 RPKM	biallelic 0.55 (40 reads), 1.43 RPKM	biallelic 0.73 (26 reads), 0.44 RPKM	No?	MEG	No?	MEG

B

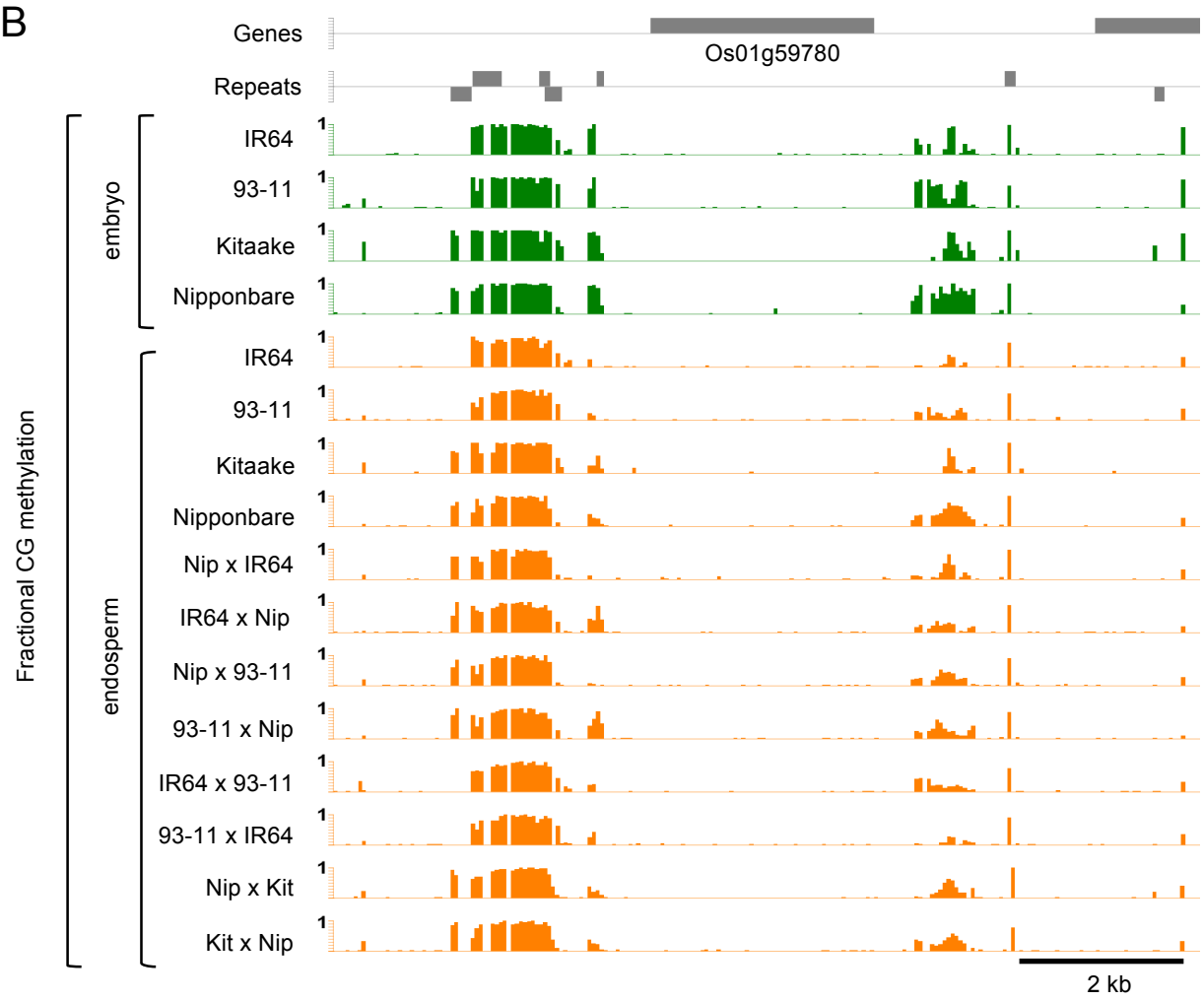


Figure 5.10 The parental bias of *Os01g59780* varies across crosses (A) but DNA methylation (B) in embryo (green) and endosperm (orange) correlates with neither apparent developmental resetting of the imprint nor potential imprinting divergence. Nip = Nipponbare; Kit = Kitaake; MEG = Maternally expressed gene; PEG = paternally expressed gene; strongly mat = strongly maternal; RPKM = reads per kilobase per million mapped reads.

Some apparent resetting may be due to contamination with seed coat or cases of imprinting divergence obscured by noise

Another instance of variation in imprinting that cannot be explained by DNA methylation and small RNA profiles is that of maternally expressed *Os01g59780* (Figure 5.10A). The gene appears to be maternally expressed in both reciprocal crosses of the Nipponbare – 93-11 set but the IR64 x Nipponbare cross, which should show a maternal bias if the Nipponbare allele is maternally expressed, shows biallelic expression. DNA methylation profiles at this locus (Figure 5.10B) do not reveal differences in methylation or hypomethylation between IR64 x Nipponbare and 93-11 x Nipponbare. Similarly, the profiles also show a lack of clear differences between cultivars. It could be that the gene is not truly imprinted and that maternal bias is a result of contamination with maternal tissues. Comparing the expression level of this gene in our samples to that in Nipponbare leaves, root, shoot, and seedling (described in Chapter IV) shows its expression in endosperm is at least 2-fold that in other tissues, though it is hard to rule out that it is more highly expressed in the seed coat than in all other tissues including endosperm.

However, there is one candidate locus at which DNA methylation profiles helped resolve possible explanations for apparent variation in imprinting. Maternally expressed *Os10g25670* (Table 5.6, Figure 5.11) initially appeared to exhibit conserved imprinting among Nipponbare, IR64 and 93-11 with developmental resetting to biallelic expression in the 93-11 x Nipponbare cross. However, DNA methylation profiles in the endosperm and embryo reveal it is likely a case of imprinting divergence combined with technical artefacts. The DNA methylation profiles suggest that alleles of the IR64 and 93-11 indica cultivars are imprinted and maternally expressed while alleles of the Nipponbare and Kitaake japonica cultivars are not. In this case, the cross with the aberrant parent-of-origin-specific bias is IR64 x Nipponbare. It is expected that this cross would be biallelic but instead it exhibits a strong maternal bias (54 out of 57 reads are maternal in origin). It is possible that this discrepancy is due to mismapped reads, poor complexity of RNA-sequencing reads or even just the statistical odds of sampling.

Table 5.6 Summary of parent-of-origin-specific expression biases at the *Os10g25670* locus in various reciprocal crosses that indicate that developmental resetting of the imprint might occur in the 93-11xNipponbare cross. Nip = Nipponbare; Kit = Kitaake; MEG = Maternally expressed gene; compl. mat = completely maternal; strongly mat = strongly maternal; RPKM = reads per kilobase per million mapped reads.

Gene locus ID	Annotated function	Fraction of maternally-sorted reads (Total number of sorted reads reads), and RPKM in crosses								Imprinted in cultivar?			
		Nip × IR64	IR64 × Nip	Nip × 93-11	93-11 × Nip	IR64 × 93-11	93-11 × IR64	Nip × Kit	Kit × Nip	IR64	93-11	Kit	Nip
Os10g25670	unknown	compl. mat 1.00 (131 read), 1.12 RPKM	strongly mat 0.95 (57 reads), 0.35 RPKM	compl. Mat 1.00 (114 reads), 0.91 RPKM	biallelic 0.71 (73 reads), 0.56 RPKM	N/A (0 reads), 0.39 RPKM	N/A (0 reads), 0.10 RPKM	N/A (0 reads), 1.75 RPKM	N/A (0 reads), 1.18 RPKM	MEG	MEG	?	MEG

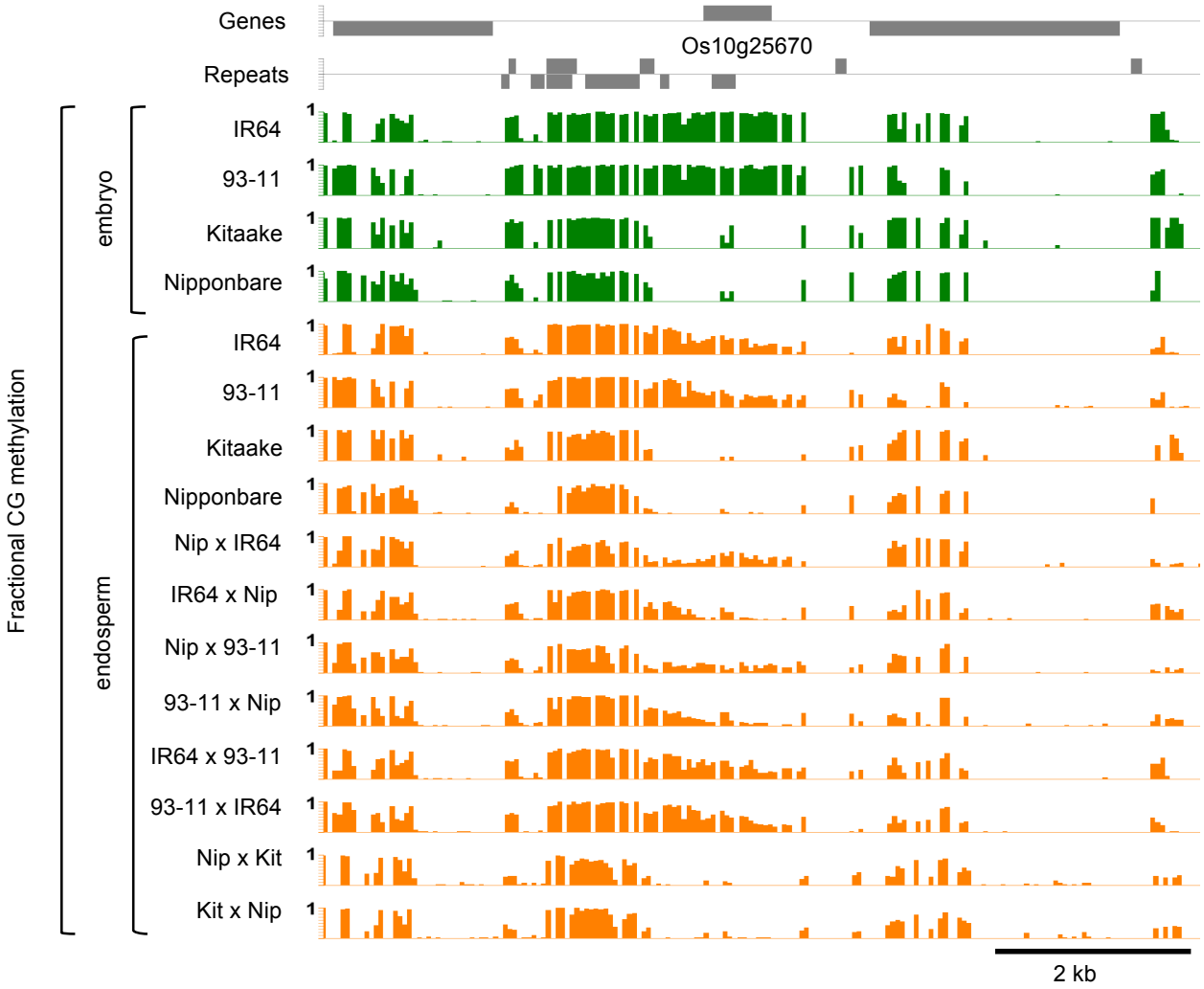


Figure 5.11 DNA methylation in embryo (green) and endosperm (orange) correlates with imprinting divergence at *Os10g25670*. Methylation and endosperm-specific hypomethylation in the gene body, transcription termini and flanking regions of the IR64 and 93-11 alleles suggest that these alleles are imprinted and maternally expressed. Nipponbare and Kitaake alleles on the other hand lack methylation at the gene locus and so are expressed from both alleles. Nip = Nipponbare; Kit = Kitaake.

Discussion

The association of epigenetic and genetic variation with imprinting divergence described in this study reinforces our previous model of imprinting regulation, which was based on genome-wide patterns of DNA methylation in imprinted genes of both *Arabidopsis thaliana* and rice (Ibarra et al., 2012; Rodrigues et al., 2013). In our model, endosperm-specific maternal DNA hypomethylation serves as a primary imprint at both maternally- and paternally expressed genes, based on the prediction that the underlying methylation at each class of imprinted gene serves a

different purpose. At paternally expressed genes, maternal-specific DNA hypomethylation in the promoter and gene body results in silencing of the maternal allele, most likely by exposing polycomb group recognition features, while the paternal allele remains expressed as it is protected from recognition by the polycomb group by DNA methylation. In contrast, at maternally expressed genes, maternal-specific DNA hypomethylation in the promoter and transcription start site activate expression from the maternal allele, while the methylated paternal allele continues to be repressed.

This pattern of enriched endosperm-specific maternal hypomethylation at promoters and gene bodies of PEGs and near the transcription start site of MEGs has also been recently found in maize (M. Zhang et al., 2014). Another recent study investigating imprinting divergence in *Arabidopsis thaliana* confirms the applicability of the model across monocots and dicots, with the authors observing that imprinting divergence at paternally expressed genes is associated with DNA methylation epialleles in the promoter and gene body (Pignatta et al., 2014). However, as they only described genetic and epigenetic variation at one divergent maternally expressed gene whose imprinting did not correlate with differential methylation, it is unclear whether the same trends we see for imprinting divergence at rice maternally expressed genes regulated by DNA hypomethylation are true in *Arabidopsis*.

In any case, the finding that maternal-specific hypomethylation is the primary imprint for most paternally expressed genes and many maternally expressed genes has interesting implications for imprinted gene evolution. It allows us to mechanistically explain the imprinting states of genes like *Os01g57890*, which appears to be maternally expressed in one cultivar and paternally expressed in the other. In the cultivar where *Os01g57890* is maternally expressed, maternal demethylation at the promoter and transcription start site appears to specifically activate the maternal allele, while in the cultivars where *Os01g57890* is paternally expressed, maternal demethylation in the promoter and gene body regions appears to specifically silence the maternal allele, probably through recruitment of the polycomb group complex. It is unclear whether the two imprinting states arose independently from a biallelically-expressed ancestral allele or whether the ancestral allele was initially imprinted in one direction and epimutation changed the direction of the bias in one of its descendant lineages. In either case, this finding lends support to the theory that both maternally and paternally expressed imprinted loci evolve from a subset of genomic loci with epigenetic features conducive to imprinting, as suggested in animal systems (Dunzinger, Nanda, Schmid, Haaf, & Zechner, 2005).

The finding that maternal-specific hypomethylation is the primary imprint for most paternally expressed genes and many maternally expressed genes may also have repercussions for how we interpret the selective pressures guiding paternal expression biases. Mammalian gametes are imprinted both in male and female lineages, with the gamete-specific deposition of methylation either preventing insulator function or silencing coding or non-coding transcripts regulated by the region (Ferguson-Smith, 2011). The observation that the only known primary 'imprint' at plant imprinted genes is a result of DNA demethylation rather than *de novo* DNA methylation and that processes that generate the imprint are specific to the maternal germline may reflect major differences in germline development between plants and animals and the constraints they place on imprinting processes.

Under parental-conflict, one could hypothesize that the maternal germline contributes to a reduction in expression of paternally expressed genes, which are expressed by default, and an increase in expression of maternally expressed genes, which are silenced by default, in order to further the interests of the maternal parent and moderate growth of the offspring. The paternal germline on the other hand is not able to directly engage in conflict, possibly due to developmental constraints on the extent of reprogramming. However, this situation would not explain cases such as *Os01g57890*, which appears to be maternally expressed in one cultivar and paternally expressed in the other. Also, assuming a passive role of the male parent in imprinting regulation, one would not expect imprinted genes to diverge at an accelerated rate compared to the rest of the genome as conditions for the Red Queen effect would not be met (Brockhurst et al., 2014). In line with this, imprinted genes in maize do not show an accelerated rate of evolution when gene sequences are compared to their orthologs in teosinte, a wild relative of maize (Waters et al., 2013). Parental conflict and the Red Queen effect may better explain imprinting in mammals, where imprinting control is less one-sided (Ferguson-Smith, 2011).

It is noteworthy that not all genes in our list of imprinted genes appear to conform to our model. The list of maternally expressed genes identified by us seems to include two distinct groups: one which shows endosperm-preferred expression compared to vegetative tissues and is associated with endosperm-specific DNA hypomethylation at the promoter and/or transcription start site, and a second group that is not endosperm-preferred and is not as highly enriched for endosperm-specific DNA hypomethylation. It is unclear what mechanisms regulate the imprinting of this second group, which is comprised of 102 of the 253 genes that are maternally expressed in at least two cultivars after filtering for possible seed coat contamination. MEGs in maize (M. Zhang et al., 2014) also seem to divide into these two groups. MEGs in *Arabidopsis* show an even greater proportion of genes that are unaffected by changes in DNA methylation (Hsieh et al., 2011), though it has been proposed that many of these genes are not truly imprinted but instead expressed in maternal tissues surrounding the endosperm.

As discussed in Chapter IV, it is possible that many of the genes in this second group are not truly maternally expressed imprinted genes, but instead the results of erroneous calls due to contamination with maternally-derived seed coat RNA. We and others (M. Zhang et al., 2014) have tried to assess endosperm-specificity by comparisons with vegetative tissue such as shoots and leaves, however, it is likely the transcriptiome and epigenome of the seed coat are significantly different from that of shoot and leaf tissues. Studies of the seed coat of *Arabidopsis thaliana* have revealed that it is a complex tissue with spatiotemporal variation in gene regulation and function (Khan, Chan, Millar, Girard, & Belmonte, 2014) and it is likely that the same is true in several other species including rice (Radchuk & Borisjuk, 2014). Seed coat development is even linked to growth of the endosperm and embryo, with two seed coat-expressed genes in rice demonstrating effects on seed size (Izawa et al., 2010; E. Wang et al., 2008). One of these genes, *Os04g33740*, which is expressed specifically in the maternal vasculature of the seed coat that delivers nutrients to the endosperm (E. Wang et al., 2008), was an initial MEG candidate in our list but was filtered out using expression data from rice leaves, shoots and seedlings (Chapter IV). It is also possible that the word 'contaminant' may be a misnomer for such transcripts as RNAs produced in maternally-derived tissues may naturally diffuse or be transported into the endosperm, where they could play key roles in development.

We also find examples of paternally expressed genes, such as *Os11g16590* and *Os06g42910*, that are not significantly enriched for maternal-specific DNA hypomethylation. The importance of maternal hypomethylation as a primary imprint regulating most paternally expressed genes was highlighted by the previously-mentioned study in maize, which also found that parent-of-origin-specific H3K27me3 was specific to the maternal genome in all assessable instances and was enriched in PEGs compared to other genes (M. Zhang et al., 2014). Similar to our expectations in rice, most paternally-biased gene expression in maize is associated with polycomb group-mediated repression of the maternal allele in endosperm and expression in non-endosperm tissues (M. Zhang et al., 2014). However, the authors of that study also report a second group of PEGs that are endosperm-specific but still have the same profile of H3K27me3 deposition on maternal but not paternal alleles. As these genes appear to have a default 'off' state, they posit that such genes undergo specific removal of H3K27me3 on paternal but not maternal alleles; however, in the absence of a control tissue that truly represents the default state of meristematic cells prior to gametogenesis, it is more plausible that endosperm-specific PEGs may also be the result of polycomb group-mediated silencing of maternal alleles induced by maternal hypomethylation.

Even though moderately paternally-biased *Os11g16590* does not correlate with endosperm-specific gene body hypomethylation, as is usual of paternally expressed genes, it possesses small regions of hypomethylation in the promoter and lies in very close proximity to the paternally expressed gene *Os11g16580*, which does display promoter and gene body hypomethylation. It is possible that epigenetic marks silencing the maternal allele of *Os11g16580* may spread to *Os11g16590*. This would be reminiscent of mammalian 'imprinting clusters' (Ferguson-Smith, 2011), where a single imprinting control region regulates imprinted expression at several genes in the cluster. However, imprinting clusters described in flowering plants are much smaller in scale than those in mammals (Luo et al., 2011; Waters et al., 2011; Wolff et al., 2011; M. Zhang et al., 2011) and our data are consistent with those of others suggesting that it is not a major mechanism of imprinting. This difference between mammalian and plant imprinting control regions may be due to the observed tendency of animal enhancers to act over a much longer range than their plant equivalents, though both clades lack a good general description of the mechanism of enhancer activity (Singer, Cox, & Liu, 2010).

In contrast to *Os11g16590*, the *Os06g42910* locus does not display any endosperm-specific DNA hypomethylation within 5 kb of the gene, though it is possible that developmental resetting at this locus might confound the analysis. As at all other candidate genes that appeared to show imprinted expression of a cultivar allele in one cross and biallelic expression in the other, we were unable to observe changes in DNA methylation and small RNA profiles that correlated with apparent resetting. It is possible that this is because the apparent variation is a result of technical artefacts rather than true developmental resetting of the imprint, or it may be due to individual cells in a population having equal contributions to the average DNA methylation profile of the population but not the average RNA expression profile of the population. The role of the abundant small RNAs present at *Os06g42910*, either in developmental resetting or in maintenance of the imprint, is unclear.

In general, most imprinted genes are not associated with imprinted small RNAs and we were unable to identify clear cases where imprinting divergence was associated with differences in small RNA expression or bias. It is also unknown whether 24-nt small RNAs detected in the endosperm play a role in directing DNA methylation to the same extent that they do in vegetative tissues, as low expression of DNA methyltransferases was observed in the young endosperm of *Arabidopsis* (Jullien et al., 2012). In fact, *Arabidopsis* seeds show an anticorrelation between the expression of genes in the RNA-dependent DNA methylation (RdDM) pathway and imprinted expression in developing endosperm: relatively high transposon activity and few RdDM gene transcripts are seen in the chalazal endosperm (Belmonte et al., 2013), where most imprinted gene expression occurs (Pignatta et al., 2014). Expression of RdDM pathway components is restored to higher levels later in *Arabidopsis* seed development, though the endosperm is reduced to a thin layer a few cells at that stage (Belmonte et al., 2013).

At the stage of endosperm used in our experiments, transcription of several small RNA and DNA methylation pathway components (DCL3a, AGO4a, AGOb, RDR2, DRM2, DCL1, MET1b and CMT2) was 2 to 20 times higher than that in seedling, leaf, root, or shoot tissues. In fact, the only pathway component expressed less in the endosperm than other tissues was DCL2b, whose function is unknown. However, transcript presence does not equate to enzymatic activity, and there is much to learn about the biogenesis and function of small RNAs in rice endosperm. Imprinting provides a great opportunity to undertake such investigations, as comparisons can be drawn between silenced and active alleles in the same environment such that cis-acting factors are teased apart from trans-acting factors.

Indeed, an understanding of mechanisms associated with imprinted expression at specific loci in mammals has shed much light on the relationship between DNA methylation, non-coding RNAs and gene expression in a developmental context (Ferguson-Smith, 2011), and it is likely that investigations of imprinting variation in plants should do the same. The general mechanism of polycomb recruitment and activation in plants is still poorly understood at this stage (Kim & Sung, 2014) and is likely to benefit from explorations of PEG imprinting divergence. In particular, information about the epigenetic factors that intiate polycomb regulation of a locus may be gleaned from an investigation of the divergence of *Os01g57890* between maternally and paternally expressed states, and further insights may be acquired by contrasting the copia retrotransposon insertions in paternally expressed *Os02g57200* and *Os11g09329*, which appear to be similar in age and yet have opposite effects on imprinted expression.

Retrotransposition events in general seem to have played a significant role in recent imprinting evolution in rice. Of our list of 19 genes with diverged imprinting among rice cultivars, two were maternally-expressed retrotransposon proteins, and all imprinting divergence associated with genetic variation (3 out of 7 assessed genes) involved retrotransposition events. Comparisons of both indica and japonica genomes of *Oryza sativa L.* to the African rice species *Oryza glaberrima* indicate that bursts of LTR retrotransposon amplification have occurred independently in each subspecies after their divergence from a common ancestor (Ma & Bennetzen, 2004). Such young LTR retrotransposition events tend to be found closer to genes and to be characterized by higher levels of DNA methylation than older LTR retrotransposition events (Vonholdt, Takuno, & Gaut, 2012).

While most LTR-captured genes in rice appear to evolve into pseudogenes, some genes remain functional and show evidence of positive selection, suggesting that LTR retrotransposition events may lead to the evolution of new gene functions (S. Y. Jiang & Ramachandran, 2013). A role for transposition events and related methylation dynamics in imprinting evolution has been previously implied by work in *Arabidopsis* (Gehring, Bubb, & Henikoff, 2009; Ibarra et al., 2012; Y. Kinoshita et al., 2007; Makarevich et al., 2008; Villar et al., 2009) and maize (Walker, 1998). It has been posited that ongoing transposition events in plant lineages such as maize and rice might provide a substantial pool of imprinted genes that selection has yet to act on (Waters et al., 2013).

It is interesting to speculate that the rate of imprinting evolution may be closely linked to general levels of transposon activity. Retrotransposition in plant lineages is thought to occur in bursts rather than steady activity (El Baidouri & Panaud, 2013), so it is possible that the rate of imprinting in plants also widely fluctuates over evolutionary time. The results of several interploidy crosses suggest that perturbations to imprinted expression of specific genes can disrupt endosperm development and act as a postzygotic hybridization barrier (Kradolfer, Wolff, Jiang, Siretskiy, & Kohler, 2013; Sekine et al., 2013), so it is possible that the misregulation of imprinted genes may provide a link between bursts of transposon-activity and speciation (Oliver, McComb, & Greene, 2013). This may also explain why it is common to see genes imprinted in one plant species not imprinted in the other, especially when the evolutionary histories of the two species include bursts of transposon activity since their divergence.

Materials and Methods

Library preparation for sequencing of small RNAs and bisulfite-treated DNA

Experiments utilized embryo and endosperm from self-fertilized seeds and cross-fertilized seeds grown and harvested at day 7 to 8 after pollination, as described in Chapters II and IV. Milky stage endosperm was pipetted out from inside the seed coat and stored either in TRIzol reagent (Ambion) for preparing small RNA libraries or 2% CTAB (Hexadecyltrimethylammonium bromide) for bisulfite sequencing libraries. Embryos were isolated after the endosperm had been collected and were washed individually through vigorous agitation in 0.5 ml of 1x Phosphate Buffered Saline solution.

Individually isolated F1 seeds were verified for heterozygosity with a PCR-based assay using either microsatellite marker RM1 or RM 72 (http://www.gramene.org/db/markers/marker_view?marker_id=24985466; McCouch et al., 2002), as described in Chapter IV. Bisulfite libraries were prepared exactly as described in Chapter II and small RNA libraries were prepared exactly as described in Chapter III. Sequencing was performed on the Illumina HiSeq 2000 platform run by the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley, generating 50 bp single-end reads for small RNA sequencing libraries and 100 bp single-end reads for bisulfite-sequencing libraries.

Allele-specific mapping and visualization of reads

All analyses were performed with either the Nipponbare rice reference genome (MSU 7.0) or with pseudo-genomes built using the Nipponbare reference and lists of identified SNPs. Pseudo-genomes were built as described in (Hsieh et al., 2011). The gene annotations used were from the MSU version 7.0 Nipponbare rice genome, while repeats were annotated using RepeatMasker with the Viridiplantae Repbase database of repetitive sequenes.

For small RNA-sequencing reads, adapter sequences were trimmed from the reads and resultant trimmed reads were sorted into size classes from 17 nt to 30 nt using custom Python scripts. Bowtie (Langmead et al., 2009) was used to independently the 24-nt size class to both maternal and paternal genomes, and custom Perl scripts (Hsieh et al., 2011) were used to sort reads as described in Chapter III. Abundances of total small RNA across the genome were calculated in reads per million (RPM).

For bisulfite-sequencing reads, alignment to both parental genomes was once again performed using Bowtie (Langmead et al., 2009), as described in Chapter II. DNA methylation of cytosines within sorted reads was calculated as described (T.-F. Hsieh et al., 2009; Zemach, McDaniel, et al., 2010) and results were visualized on SignalMap2 (Nimblegen) software.

Sanger sequencing of loci with diverged imprinting

Primers were designed to amplify overlapping regions of the Nipponbare genome sequence that were between 823 and 2371 bp in length, and 1678 bp on average. Primers were 25 to 35 long, with GC content between 42 and 60% and melting temperature between 68 and 77°C. Initial PCRs were done with ExTaq DNA polymerase (Takara Bio) but were repeated with Phusion High-Fidelity DNA polymerase (NEB) if poor quality amplification or sequencing was obtained with ExTaq. ExTaq PCR conditions were as follows: 1.25 U of ExTaq DNA polymerase, 5 µl of 10X Extaq reaction buffer, 200 µM dNTPs, 0.2 µM each primer, and 5 ng genomic DNA (50 µl final). Two-step PCR reactions were carried out as follows: 98℃ 1 min, then 30 cycles of 98℃ 10 sec and 68℃ 6 min, and a final extension of 72℃ 15 min. Phusion PCR conditions were as follows: 0.4 U of Phusion DNA polymerase, 4 µl of 15X Phusion GC buffer, 200 µM dNTPs, 0.5 µM each primer, 2M Betaine (Sigma-Aldrich) and 5 ng genomic DNA (20 µl final). Two-step PCR reactions were carried out as follows: 98℃ 3 min, then 30 cycles of 98℃ 10 sec and 68℃ 6 min, and a final extension of 72℃ 10 min. Annealing temperatures were adjusted as needed to increase specificity. PCR products were treated with Shrimp Alkaline Phosphatase (NEB) and Exonuclease I (NEB) and then Sanger sequenced with forward and reverse primers at the DNA Sequencing Facility at Barker Hall, UC Berkeley.

Chapter VI

Concluding remarks

When I began this project, about a dozen imprinted genes had been identified in plants—handfuls in any single plant species. Similarities and differences in mechanisms of imprinting at maternally and paternally expressed genes were poorly understood, and it was unclear what evolutionary forces and epigenetic and genetic mechanisms were guiding imprinting evolution and divergence. It was however known that certain imprinted genes in *Arabidopsis* played vital functions in endosperm development and that perturbations in imprinted expression resulted in abnormal endosperm development, including changes in size or even growth arrest and seed abortion. For handfuls of genes, maternal specific DNA demethylation and/or histone modification by the polycomb group contributed to the differential activation and silencing of maternal and paternal alleles, but it was unknown whether these regulatory paradigms were applicable to the bulk of plant imprinted genes.

The field has advanced a great deal during the last five years, with improvements in high throughput sequencing technology and the availability of reference genome sequence from multiple plant species leading to comprehensive genome-scale investigations of imprinted expression and epigenetic marks that correlate with it. The wealth of data has allowed several comparisons of imprinting targets and mechanisms, both within and among species, so that the evolutionary forces guiding imprinted expression may be better understood and the functional significance of imprinting elucidated. Similarities in imprinting mechanisms and functional enrichment of imprinted genes across monocots and dictots suggest that common forces guide the evolution of imprinting in all plant species. However, these forces are likely to differ from those acting to drive imprinting evolution in mammals, given differences in developmental program and the unique triploid state of plant endosperm. It is possible that the widely accepted parental-conflict hypothesis of genomic imprinting, though highly applicable to mammalian imprinting, might not be the greatest evolutionary driver of plant imprinting in recent plant lineages. It is likely that the imprinting of different genes is shaped by different evolutionary forces or combinations of evolutionary forces, and that the dosage hypothesis may provide a better general explanation for imprinting.

In summary, I have generated models of imprinting regulation that appear to apply to most paternally expressed genes and a large number of maternally expressed genes in rice, and note the work of others in showing that these models extend to two other species across flowering plants. I identify LTR retrotransposition as a mechanism that has contributed to imprinting divergence over short evolutionary timescales in rice. My work was also the first to show both maternal and paternal biases in seed small RNAs and to identify the correlation between imprinted expression of small RNAs and that of genes, though more work remains to be done in characterizing the role these small RNAs play in maintenance or potential resetting of the imprint. My findings make valuable contributions to the understanding of mechanisms of epigenetic regulation, seed development, hybrid incompatibilities, and plant evolution, and have the potential to contribute to improvements in breeding strategies for cereal crops that feed much of the world.

References

Adalsteinsson, B. T., & Ferguson-Smith, A. C. (2014). Epigenetic control of the genome-lessons from genomic imprinting. *Genes (Basel), 5*(3), 635-655. doi: 10.3390/genes5030635

Adams, I. R., & McLaren, A. (2002). Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis. *Development, 129*(5), 1155-1164.

Adenot, P. G., Mercier, Y., Renard, J. P., & Thompson, E. M. (1997). Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. *Development, 124*(22), 4615-4625.

Agarwal, P., Kapoor, S., & Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays, 33*(3), 189-202. doi: 10.1002/bies.201000107

Alvarez-Buylla, E. R., Benitez, M., Corvera-Poire, A., Chaos Cador, A., de Folter, S., Gamboa de Buen, A., . . . Sanchez-Corrales, Y. E. (2010). Flower development. *Arabidopsis Book, 8*, e0127. doi: 10.1199/tab.0127

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. *Genome Biol, 8*(9), 227. doi: 10.1186/gb-2007-8-9-227

Barlow, D. P. (1993). Methylation and imprinting: from host defense to gene regulation? *Science, 260*(5106), 309-310.

Barlow, D. P. (1994). Imprinting: a gamete's point of view. *Trends Genet, 10*(6), 194-199.

Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K., & Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature, 349*(6304), 84-87. doi: 10.1038/349084a0

Baroux, C., Pecinka, A., Fuchs, J., Schubert, I., & Grossniklaus, U. (2007). The triploid endosperm genome of Arabidopsis adopts a peculiar, parental-dosage-dependent chromatin organization. *Plant Cell, 19*(6), 1782-1794. doi: 10.1105/tpc.106.046235

Bartolomei, M. S., & Ferguson-Smith, A. C. (2011). Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol, 3*(7). doi: 10.1101/cshperspect.a002592

Bartolomei, M. S., Webber, A. L., Brunkow, M. E., & Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev, 7*(9), 1663-1673.

Bartolomei, M. S., Zemel, S., & Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature, 351*(6322), 153-155. doi: 10.1038/351153a0

Bateson, W. (1926). Segregation: Being the Joseph Leidy memorial lecture of the University of Pennsylvania, 1922. *Journal of Genetics, 16*(2), 201-235.

Bauer, M. J., & Fischer, R. L. (2011). Genome demethylation and imprinting in the endosperm. *Curr Opin Plant Biol, 14*(2), 162-167. doi: S1369-5266(11)00011-2 [pii], 10.1016/j.pbi.2011.02.006

Baurle, I., & Dean, C. (2006). The timing of developmental transitions in plants. *Cell, 125*(4), 655-664. doi: 10.1016/j.cell.2006.05.005

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature, 405*(6785), 482-485. doi: 10.1038/35013100

Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., . . . Harada, J. J. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A, 110*(5), E435-444. doi: 10.1073/pnas.1222061110

Berger, F., & Twell, D. (2011). Germline specification and function in plants. *Annu Rev Plant Biol, 62*, 461-484. doi: 10.1146/annurev-arplant-042110-103824

Birchler, J. A. (2014). Interploidy hybridization barrier of endosperm as a dosage interaction. *Front Plant Sci, 5*, 281. doi: 10.3389/fpls.2014.00281

Bleckmann, A., Alter, S., & Dresselhaus, T. (2014). The beginning of a seed: regulatory mechanisms of double fertilization. *Front Plant Sci, 5*, 452. doi: 10.3389/fpls.2014.00452

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature, 431*(7004), 96-99. doi: 10.1038/nature02886

Bourc'his, D., & Bestor, T. H. (2006). Origins of extreme sexual dimorphism in genomic imprinting. *Cytogenet Genome Res, 113*(1-4), 36-40. doi: 10.1159/000090813

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science, 294*(5551), 2536-2539. doi: 10.1126/science.1065848

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. *RNA, 19*(4), 429-442. doi: 10.1261/rna.037598.112

Brockhurst, M. A., Chapman, T., King, K. C., Mank, J. E., Paterson, S., & Hurst, G. D. (2014). Running with the Red Queen: the role of biotic conflicts in evolution. *Proc Biol Sci, 281*(1797). doi: 10.1098/rspb.2014.1382

Burkhardt, P. K., Beyer, P., Wunn, J., Kloti, A., Armstrong, G. A., Schledz, M., . . . Potrykus, I. (1997). Transgenic rice (Oryza sativa) endosperm expressing daffodil (Narcissus pseudonarcissus) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. *Plant J, 11*(5), 1071-1078.

Calarco, J. P., Borges, F., Donoghue, M. T., Van Ex, F., Jullien, P. E., Lopes, T., . . . Martienssen, R. A. (2012). Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell, 151*(1), 194-205. doi: S0092-8674(12)01066-5 [pii], 10.1016/j.cell.2012.09.001

Chaw, S. M., Chang, C. C., Chen, H. L., & Li, W. H. (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol, 58*(4), 424-441. doi: 10.1007/s00239-003-2564-9

Chen, M., Presting, G., Barbazuk, W. B., Goicoechea, J. L., Blackmon, B., Fang, G., . . . Wing, R. A. (2002). An integrated physical and genetic map of the rice genome. *Plant Cell, 14*(3), 537-545.

Chodavarapu, R. K., Feng, S., Ding, B., Simon, S. A., Lopez, D., Jia, Y., . . . Pellegrini, M. (2012). Transcriptome and methylome interactions in rice hybrids. *Proc Natl Acad Sci U S A, 109*(30), 12040-12045. doi: 10.1073/pnas.1209297109

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., . . . Fischer, R. L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell, 110*(1), 33-42. doi: S0092867402008073 [pii]

Chotalia, M., Smallwood, S. A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., . . . Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev, 23*(1), 105-117. doi: 10.1101/gad.495809

Coffman, W. R., & Hargrove, T. R. (1989). Modern rice varieties as a possible factor in production variability. In J. R. Anderson & P. B. R. Hazell (Eds.), *Variability in grain yields: implications for agricultural research and policy in developing countries* (pp. 133-146): International Food Policy Research Institute.

Costa, L. M., Yuan, J., Rouster, J., Paul, W., Dickinson, H., & Gutierrez-Marcos, J. F. (2012). Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr Biol, 22*(2), 160-165. doi: 10.1016/j.cub.2011.11.059

Couvillion, M. T., Lee, S. R., Hogstad, B., Malone, C. D., Tonkin, L. A., Sachidanandam, R., . . . Collins, K. (2009). Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila*. *Genes Dev, 23*(17), 2016-2032. doi: 10.1101/gad.1821209

Crespi, B., & Semeniuk, C. (2004). Parent-offspring conflict in the evolution of vertebrate reproductive mode. *Am Nat, 163*(5), 635-653. doi: 10.1086/382734

Csorba, T., Questa, J. I., Sun, Q., & Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc Natl Acad Sci U S A, 111*(45), 16160-16165. doi: 10.1073/pnas.1419030111

DeChiara, T. M., Robertson, E. J., & Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell, 64*(4), 849-859.

Del Toro-De Leon, G., Garcia-Aguilar, M., & Gillmor, C. S. (2014). Non-equivalent contributions of maternal and paternal genomes to early plant embryogenesis. *Nature, 514*(7524), 624-627. doi: 10.1038/nature13620

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., & Jacobsen, S. E. (2012). Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. *PLoS Genet, 8*(11), e1003062. doi: 10.1371/journal.pgen.1003062

Dilkes, B. P., & Comai, L. (2004). A differential dosage hypothesis for parental effects in seed development. *Plant Cell, 16*(12), 3174-3180. doi: 10.1105/tpc.104.161230

Drews, G. N., & Koltunow, A. M. (2011). The female gametophyte. *Arabidopsis Book, 9*, e0155. doi: 10.1199/tab.0155

Du, M., Luo, M., Zhang, R., Finnegan, E. J., & Koltunow, A. M. (2014). Imprinting in rice: the role of DNA and histone methylation in modulating parent-of-origin specific expression and determining transcript start sites. *Plant J, 79*(2), 232-242. doi: 10.1111/tpj.12553

Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res, 38*(Web Server issue), W64-70. doi: 10.1093/nar/gkq310

Dunzinger, U., Nanda, I., Schmid, M., Haaf, T., & Zechner, U. (2005). Chicken orthologues of mammalian imprinted genes are clustered on macrochromosomes and replicate asynchronously. *Trends Genet, 21*(9), 488-492. doi: 10.1016/j.tig.2005.07.004

El Baidouri, M., & Panaud, O. (2013). Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol Evol, 5*(5), 954-965. doi: 10.1093/gbe/evt025

Erhardt, S., Su, I. H., Schneider, R., Barton, S., Bannister, A. J., Perez-Burgos, L., . . . Surani, M. A. (2003). Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development, 130*(18), 4235-4248.

Ewen, K. A., & Koopman, P. (2010). Mouse germ cell development: from specification to sex determination. *Mol Cell Endocrinol, 323*(1), 76-93. doi: 10.1016/j.mce.2009.12.013

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet, 12*(8), 565-575. doi: 10.1038/nrg3032

Ferguson-Smith, A. C., Sasaki, H., Cattanach, B. M., & Surani, M. A. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature, 362*(6422), 751-755. doi: 10.1038/362751a0

Ferron, S. R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., . . . Ferguson-Smith, A. C. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature, 475*(7356), 381-385. doi: 10.1038/nature10229

Fitzpatrick, G. V., Soloway, P. D., & Higgins, M. J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet, 32*(3), 426-431. doi: 10.1038/ng988

Frost, J. M., & Moore, G. E. (2010). The importance of imprinting in the human placenta. *PLoS Genet, 6*, e1001015. doi: 10.1371/journal.pgen.1001015

Gehring, M. (2013). Genomic imprinting: insights from plants. *Annu Rev Genet, 47*, 187-208. doi: 10.1146/annurev-genet-110711-155527

Gehring, M., Bubb, K. L., & Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science, 324*(5933), 1447-1451. doi: 324/5933/1447 [pii], 10.1126/science.1171609

Gehring, M., & Henikoff, S. (2007). DNA methylation dynamics in plant genomes. *Biochim Biophys Acta, 1769*(5-6), 276-286. doi: 10.1016/j.bbaexp.2007.01.009

Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., . . . Fischer, R. L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell, 124*(3), 495-506. doi: S0092-8674(06)00068-7 [pii], 10.1016/j.cell.2005.12.034

Gehring, M., Missirian, V., & Henikoff, S. (2011). Genomic Analysis of Parent-of-Origin Allelic Expression in Arabidopsis thaliana Seeds. *PLoS One, 6*(8), e23687. doi: 10.1371/journal.pone.0023687

Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A., & Gagliano, W. B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science, 280*(5362), 446-450.

Guibert, S., Forne, T., & Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res, 22*(4), 633-641. doi: 10.1101/gr.130997.111

Gutierrez-Marcos, J. F., Costa, L. M., Dal Pra, M., Scholten, S., Kranz, E., Perez, P., & Dickinson, H. G. (2006). Epigenetic asymmetry of imprinted genes in plant gametes. *Nat Genet, 38*(8), 876-878. doi: 10.1038/ng1828

Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., & Surani, M. A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science, 339*(6118), 448-452. doi: 10.1126/science.1229277

Haig, D. (2014). Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity (Edinb), 113*(2), 96-103. doi: 10.1038/hdy.2013.97

Haig, D., & Graham, C. (1991). Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell, 64*(6), 1045-1046.

Haig, D., & Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. *The American Naturalist, 134*(1), 147-155.

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature, 405*(6785), 486-489. doi: 10.1038/35013106

He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L. L., Manzhong, . . . Deng, X.-W. (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *The Plant Cell, 22*, 17-33.

He, X. J., Chen, T., & Zhu, J. K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Res, 21*(3), 442-465. doi: cr201123 [pii], 10.1038/cr.2011.23

Henckel, A., Chebli, K., Kota, S. K., Arnaud, P., & Feil, R. (2012). Transcription and histone methylation changes correlate with imprint acquisition in male germ cells. *EMBO J, 31*(3), 606-615. doi: 10.1038/emboj.2011.425

Hermon, P., Srilunchang, K. O., Zou, J., Dresselhaus, T., & Danilevskaya, O. N. (2007). Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. *Plant Mol Biol, 64*(4), 387-395. doi: 10.1007/s11103-007-9160-0

Horsthemke, B., & Wagstaff, J. (2008). Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am J Med Genet A, 146A*(16), 2041-2052. doi: 10.1002/ajmg.a.32364

Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., & Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. *Science, 324*(5933), 1451-1454. doi: 324/5933/1451 [pii], 10.1126/science.1172417

Hsieh, T. F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., . . . Fischer, R. L. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci U S A, 108*(5), 1755-1762. doi: 1019273108 [pii], 10.1073/pnas.1019273108

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., . . . Han, B. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res, 19*(6), 1068-1076. doi: gr.089516.108 [pii], 10.1101/gr.089516.108

Huh, J. H., Bauer, M. J., Hsieh, T. F., & Fischer, R. L. (2008). Cellular programming of plant gene imprinting. *Cell, 132*(5), 735-744. doi: S0092-8674(08)00266-3 [pii], 10.1016/j.cell.2008.02.018

Ibarra, C. A., Feng, X., Schoft, V. K., Hsieh, T. F., Uzawa, R., Rodrigues, J. A., . . . Zilberman, D. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science, 337*(6100), 1360-1364. doi: 337/6100/1360 [pii], 10.1126/science.1224839

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature, 436*(7052), 793-800. doi: 10.1038/nature03895

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., & Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncitial nuclear division. *The Plant Journal*.

Izawa, Y., Takayanagi, Y., Inaba, N., Abe, Y., Minami, M., Fujisawa, Y., . . . Iwasaki, Y. (2010). Function and expression pattern of the alpha subunit of the heterotrimeric G protein in rice. *Plant Cell Physiol, 51*(2), 271-281. doi: 10.1093/pcp/pcp186

Jahnke, S., & Scholten, S. (2009). Epigenetic resetting of a gene imprinted in plant embryos. *Curr Biol, 19*(19), 1677-1681. doi: S0960-9822(09)01630-3 [pii], 10.1016/j.cub.2009.08.053

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature, 449*(7159), 248-251. doi: 10.1038/nature06146

Jiang, H., & Kohler, C. (2012). Evolution, function, and regulation of genomic imprinting in plant seed development. *J Exp Bot, 63*(13), 4713-4722. doi: ers145 [pii], 10.1093/jxb/ers145

Jiang, S. Y., & Ramachandran, S. (2013). Genome-wide survey and comparative analysis of LTR retrotransposons and their captured genes in rice and sorghum. *PLoS One, 8*(7), e71118. doi: 10.1371/journal.pone.0071118

Johnson, M. H., & Selwood, L. (1996). Nomenclature of early development in mammals. *Reprod Fertil Dev, 8*(4), 759-764.

Jullien, P. E., Katz, A., Oliva, M., Ohad, N., & Berger, F. (2006). Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. *Curr Biol, 16*(5), 486-492. doi: 10.1016/j.cub.2006.01.020

Jullien, P. E., Susaki, D., Yelagandula, R., Higashiyama, T., & Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. *Curr Biol, 22*(19), 1825-1830. doi: 10.1016/j.cub.2012.07.061

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., & Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J, 32*(3), 340-353. doi: 10.1038/emboj.2012.331

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., . . . Sasaki, H. (2007). Role of the Dnmt3 family in *de novo* methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet, 16*(19), 2272-2280. doi: 10.1093/hmg/ddm179

Keller, C., & Buhler, M. (2013). Chromatin-associated ncRNA activities. *Chromosome Res, 21*(6-7), 627-641. doi: 10.1007/s10577-013-9390-8

Kermicle, J. L. (1970). Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics, 66*(1), 69-85.

Khan, D., Chan, A., Millar, J. L., Girard, I. J., & Belmonte, M. F. (2014). Predicting transcriptional circuitry underlying seed coat development. *Plant Sci, 223*, 146-152. doi: 10.1016/j.plantsci.2014.03.016

Kim, D. H., & Sung, S. (2014). Polycomb-Mediated Gene Silencing in Arabidopsis thaliana. *Mol Cells*. doi: 10.14348/molcells.2014.0249

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., . . . Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science, 303*(5657), 521-523. doi: 10.1126/science.1089835, 1089835 [pii]

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W. J., Koornneef, M., & Kakutani, T. (2007). Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J, 49*(1), 38-45. doi: TPJ2936 [pii], 10.1111/j.1365-313X.2006.02936.x

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., et al. (1999). Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. *Proc Natl Acad Sci U S A, 96*(7), 4186-4191.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., . . . Kono, T. (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res, 23*(4), 616-627. doi: 10.1101/gr.148023.112

Kohler, C., Page, D. R., Gagliardini, V., & Grossniklaus, U. (2005). The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat Genet, 37*(1), 28-30. doi: ng1495 [pii], 10.1038/ng1495

Kohler, C., & Weinhofer-Molisch, I. (2009). Mechanisms and evolution of genomic imprinting in plants. *Heredity*. doi: hdy2009176 [pii], 10.1038/hdy.2009.176

Kradolfer, D., Wolff, P., Jiang, H., Siretskiy, A., & Kohler, C. (2013). An imprinted gene underlies postzygotic reproductive isolation in Arabidopsis thaliana. *Dev Cell, 26*(5), 525-535. doi: 10.1016/j.devcel.2013.08.006

Kuzin, A. B., Lyubomirskaya, N. V., Khudaibergenova, B. M., Ilyin, Y. V., & Kim, A. I. (1994). Precise excision of the retrotransposon gypsy from the forked and cut loci in a genetically unstable D. melanogaster strain. *Nucleic Acids Res, 22*(22), 4641-4645.

Lafon-Placette, C., & Kohler, C. (2014). Embryo and endosperm, partners in seed development. *Curr Opin Plant Biol, 17*, 64-69. doi: 10.1016/j.pbi.2013.11.008

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*(3), R25.

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet, 11*(3), 204-220. doi: 10.1038/nrg2719

Lehtonen, J., & Parker, G. A. (2014). Gamete competition, gamete limitation, and the evolution of the two sexes. *Mol Hum Reprod, 20*(12), 1161-1168. doi: 10.1093/molehr/gau068

Li, H., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res, 18*, 1851-1858.

Li, J., & Berger, F. (2012). Endosperm: food for humankind and fodder for scientific discoveries. *New Phytol, 195*(2), 290-305. doi: 10.1111/j.1469-8137.2012.04182.x

Liu, H., Kim, J. M., & Aoki, F. (2004). Regulation of histone H3 lysine 9 methylation in oocytes and early pre-implantation embryos. *Development, 131*(10), 2269-2280. doi: 10.1242/dev.01116

Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., & Kuang, H. (2012). Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. *Mol Biol Evol, 29*(3), 1005-1017. doi: msr282 [pii], 10.1093/molbev/msr282

Lu, J., Zhang, C., Baulcombe, D. C., & Chen, Z. J. (2012). Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. *Proc Natl Acad Sci U S A, 109*(14), 5529-5534. doi: 1203094109 [pii], 10.1073/pnas.1203094109

Lu, T., Zhu, C., Lu, G., Guo, Y., Zhou, Y., Zhang, Z., . . . Han, B. (2012). Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice. *BMC Genomics, 13*, 721. doi: 10.1186/1471-2164-13-721

Lucifero, D., Mann, M. R., Bartolomei, M. S., & Trasler, J. M. (2004). Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum Mol Genet, 13*(8), 839-849. doi: 10.1093/hmg/ddh104

Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J., & Chaudhury, A. (2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proc Natl Acad Sci U S A, 97*(19), 10637-10642. doi: 10.1073/pnas.170292997

Luo, M., Taylor, J. M., Spriggs, A., Zhang, H., Wu, X., Russell, S., . . . Koltunow, A. (2011). A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet, 7*(6), e1002125. doi: 10.1371/journal.pgen.1002125

Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., Smrzka, O. W., Wutz, A., . . . Barlow, D. P. (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat Genet, 25*(1), 19-21. doi: 10.1038/75546

Ma, J., & Bennetzen, J. L. (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci U S A, 101*(34), 12404-12410. doi: 10.1073/pnas.0403715101, 0403715101 [pii]

MacDonald, W. A., & Mann, M. R. (2014). Epigenetic regulation of genomic imprinting from germ line to preimplantation. *Mol Reprod Dev, 81*(2), 126-140. doi: 10.1002/mrd.22220

Makarevich, G., Villar, C. B., Erilova, A., & Kohler, C. (2008). Mechanism of PHERES1 imprinting in Arabidopsis. *J Cell Sci, 121*(Pt 6), 906-912. doi: jcs.023077 [pii], 10.1242/jcs.023077

Matvienko, M., Kozik, A., Froenicke, L., Lavelle, D., Martineau, B., Perroud, B., et al. (2013). Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. *PLoS One, 8*(2), e55913. doi: 10.1371/journal.pone.0055913

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature, 403*(6769), 501-502. doi: 10.1038/35000654

McCouch, S. R., Teytelman, L., Xu, Y., Lobos, K. B., Clare, K., Walton, M., . . . Stein, L. (2002). Development and mapping of 2240 new SSR markers for rice (Oryza sativa L.). *DNA Res, 9*(6), 199-207.

McEwen, K. R., & Ferguson-Smith, A. C. (2010). Distinguishing epigenetic marks of developmental and imprinting regulation. *Epigenetics Chromatin, 3*(1), 2. doi: 10.1186/1756-8935-3-2

McKeown, P. C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M. W., Donoghue, M. T., . . . Spillane, C. (2011). Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds. *BMC Plant Biol, 11*, 113. doi: 1471-2229-11-113 [pii], 10.1186/1471-2229-11-113

Mendel, G. (1901). Experiments on plant hybridization. *Journal of the Royal Horticultural Society of London, 26*, 1-32.

Messerschmidt, D. M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., & Knowles, B. B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science, 335*(6075), 1499-1502. doi: 10.1126/science.1216154

Morales-Ruiz, T., Ortega-Galisteo, A. P., Ponferrada-Marin, M. I., Martinez-Macias, M. I., Ariza, R. R., & Roldan-Arjona, T. (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-

methylcytosine DNA glycosylases. *Proc Natl Acad Sci U S A, 103*(18), 6853-6858. doi: 0601109103 [pii], 10.1073/pnas.0601109103

Morison, I. M., & Reeve, A. E. (1998). A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Hum Mol Genet, 7*(10), 1599-1609. doi: ddb178 [pii]

Mosher, R. A., Melnyk, C. W., Kelly, K. A., Dunn, R. M., Studholme, D. J., & Baulcombe, D. C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature, 460*(7252), 283-286. doi: nature08084 [pii], 10.1038/nature08084

Mosher, R. A., Tan, E. H., Shin, J., Fischer, R. L., Pikaard, C. S., & Baulcombe, D. C. (2011). An atypical epigenetic mechanism affects uniparental expression of Pol IV-dependent siRNAs. *PLOS ONE, 6*(10), e25756. doi: 10.1371/journal.pone.0025756

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R., & Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science, 322*(5908), 1717-1720. doi: 10.1126/science.1163802

Nasu, S., Suzuki, J., Ohta, R., Hasegawa, K., Yui, R., Kitazawa, N., . . . Minobe, Y. (2002). Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (Oryza sativa, Oryza rufipogon) and establishment of SNP markers. *DNA Res, 9*(5), 163-171.

Niklas, K. J. (1997). *The evolutionary biology of plants*. Chicago: University of Chicago Press.

Oliver, K. R., McComb, J. A., & Greene, W. K. (2013). Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol Evol, 5*(10), 1886-1901. doi: 10.1093/gbe/evt141

Ong, C. T., & Corces, V. G. (2009). Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme. *J Biol, 8*(8), 73. doi: 10.1186/jbiol165

Ong, C. T., & Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet, 15*(4), 234-246. doi: 10.1038/nrg3663

Ono, A., Yamaguchi, K., Fukada-Tanaka, S., Terada, R., Mitsui, T., & Iida, S. (2012). A null mutation of ROS1a for DNA demethylation in rice is not transmittable to progeny. *Plant J, 71*(4), 564-574. doi: 10.1111/j.1365-313X.2012.05009.x

Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., . . . Ishino, F. (2006). Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat Genet, 38*(1), 101-106. doi: 10.1038/ng1699

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., . . . Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr Biol, 10*(8), 475-478.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., . . . Buell, C. R. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res, 35*(Database issue), D883-887. doi: gkl976 [pii], 10.1093/nar/gkl976

Patten, M. M., Ross, L., Curley, J. P., Queller, D. C., Bonduriansky, R., & Wolf, J. B. (2014). The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity (Edinb), 113*(2), 119-128. doi: 10.1038/hdy.2014.29

Penterman, J., Zilberman, D., Huh, J. H., Ballinger, T., Henikoff, S., & Fischer, R. L. (2007). DNA demethylation in the Arabidopsis genome. *Proc Natl Acad Sci U S A, 104*(16), 6752-6757. doi: 0701861104 [pii], 10.1073/pnas.0701861104

Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W., & Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife, 3*, e03198. doi: 10.7554/eLife.03198

Pinney, S. E. (2014). Mammalian Non-CpG Methylation: Stem Cells and Beyond. *Biology (Basel), 3*(4), 739-751. doi: 10.3390/biology3040739

Pires, N. D., & Grossniklaus, U. (2014). Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000Prime Rep, 6*, 63. doi: 10.12703/P6-63

Proudhon, C., Duffie, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., . . . Bourc'his, D. (2012). Protection against *de novo* methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Mol Cell, 47*(6), 909-920. doi: 10.1016/j.molcel.2012.07.010

Qiu, Y., Liu, S. L., & Adams, K. L. (2014). Frequent changes in expression profile and accelerated sequence evolution of duplicated imprinted genes in arabidopsis. *Genome Biol Evol, 6*(7), 1830-1842.

Radchuk, V., & Borisjuk, L. (2014). Physical, metabolic and developmental functions of the seed coat. *Front Plant Sci, 5*, 510. doi: 10.3389/fpls.2014.00510

Raissig, M. T., Baroux, C., & Grossniklaus, U. (2011). Regulation and flexibility of genomic imprinting during seed development. *Plant Cell, 23*(1), 16-26. doi: 10.1105/tpc.110.081018

Renfree, M. B., Suzuki, S., & Kaneko-Ishino, T. (2013). The origin and evolution of genomic imprinting and viviparity in mammals. *Philos Trans R Soc Lond B Biol Sci, 368*(1609), 20120151. doi: rstb.2012.0151 [pii], 10.1098/rstb.2012.0151

Rodrigues, J. A., Ruan, R., Nishimura, T., Sharma, M. K., Sharma, R., Ronald, P. C., . . . Zilberman, D. (2013). Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci U S A, 110*(19), 7934-7939. doi: 1306164110 [pii], 10.1073/pnas.1306164110

Sabelli, P. A., & Larkins, B. A. (2009). The development of endosperm in grasses. *Plant Physiol, 149*(1), 14-26. doi: 149/1/14 [pii], 10.1104/pp.108.129437

Saitou, M., Kagiwada, S., & Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development, 139*(1), 15-31. doi: 10.1242/dev.050849

Santos, F., Peters, A. H., Otte, A. P., Reik, W., & Dean, W. (2005). Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev Biol, 280*(1), 225-236. doi: 10.1016/j.ydbio.2005.01.025

Saze, H., Tsugane, K., Kanno, T., & Nishimura, T. (2012). DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation. *Plant Cell Physiol, 53*(5), 766-784. doi: 10.1093/pcp/pcs008

Schatlowski, N., & Kohler, C. (2012). Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. *J Exp Bot, 63*(17), 6059-6067. doi: 10.1093/jxb/ers288

Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., . . . Tamaru, H. (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. *Proc Natl Acad Sci U S A, 108*(19), 8042-8047. doi: 1105117108 [pii], 10.1073/pnas.1105117108

Schrader, F. (1921). The chromosomes of *Pseudococcus nipae*. *The Biological Bulletin, 40*(5), 259-269.

Scott, R. J., Spielman, M., Bailey, J., & Dickinson, H. G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. *Development, 125*(17), 3329-3341.

Secco, D., Jabnoune, M., Walker, H., Shou, H., Wu, P., Poirier, Y., & Whelan, J. (2013). Spatio-temporal transcript profiling of rice roots and shoots in response to phosphate starvation and recovery. *Plant Cell, 25*(11), 4285-4304. doi: 10.1105/tpc.113.117325

Sekine, D., Ohnishi, T., Furuumi, H., Ono, A., Yamada, T., Kurata, N., & Kinoshita, T. (2013). Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant J, 76*(5), 792-799. doi: 10.1111/tpj.12333

Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., . . . Ishino, F. (2008). Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nat Genet, 40*(2), 243-248. doi: 10.1038/ng.2007.51

Singer, S. D., Cox, K. D., & Liu, Z. (2010). Both the constitutive cauliflower mosaic virus 35S and tissue-specific AGAMOUS enhancers activate transcription autonomously in Arabidopsis thaliana. *Plant Mol Biol, 74*(3), 293-305. doi: 10.1007/s11103-010-9673-9

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J. D., Feijo, J. A., & Martienssen, R. A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell, 136*(3), 461-472. doi: S0092-8674(08)01644-9 [pii], 10.1016/j.cell.2008.12.038

Smyth, D. R., Bowman, J. L., & Meyerowitz, E. M. (1990). Early flower development in Arabidopsis. *Plant Cell, 2*(8), 755-767. doi: 10.1105/tpc.2.8.755

Spillane, C., Schmid, K. J., Laoueille-Duprat, S., Pien, S., Escobar-Restrepo, J. M., Baroux, C., . . . Grossniklaus, U. (2007). Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature, 448*(7151), 349-352. doi: 10.1038/nature05984

Stoger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H., & Barlow, D. P. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell, 73*(1), 61-71.

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet, 9*(6), 465-476. doi: 10.1038/nrg2341

Takahashi, N., Kobayashi, R., & Kono, T. (2010). Restoration of *Dlk1* and *Rtl1* is necessary but insufficient to rescue lethality in intergenic differentially methylated region (IG-DMR)-deficient mice. *J Biol Chem, 285*(34), 26121-26125. doi: 10.1074/jbc.M109.075325

Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J. L., Jackson, S. A., . . . Ma, J. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res, 19*(12), 2221-2230. doi: 10.1101/gr.083899.108

Tomizawa, S., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G., & Sasaki, H. (2011). Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development, 138*(5), 811-820. doi: 10.1242/dev.061416

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics, 25*(9), 1105-1111. doi: 10.1093/bioinformatics/btp120

Villar, C. B., Erilova, A., Makarevich, G., Trosch, R., & Kohler, C. (2009). Control of PHERES1 imprinting in Arabidopsis by direct tandem repeats. *Mol Plant, 2*(4), 654-660. doi: ssp014 [pii], 10.1093/mp/ssp014

Vonholdt, B. M., Takuno, S., & Gaut, B. S. (2012). Recent retrotransposon insertions are methylated and phylogenetically clustered in japonica rice (Oryza sativa spp. japonica). *Mol Biol Evol, 29*(10), 3193-3203. doi: 10.1093/molbev/mss129

Vu, T. M., Nakamura, M., Calarco, J. P., Susaki, D., Lim, P. Q., Kinoshita, T., . . . Berger, F. (2013). RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. *Development, 140*(14), 2953-2960. doi: 10.1242/dev.092981

Walker, E. L. (1998). Paramutation of the r1 locus of maize is associated with increased cytosine methylation. *Genetics, 148*(4), 1973-1981.

Wang L., Si Y., Dedow L. K., Shao Y., Liu P., Brutnell T. P. (2011). A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PloS one*. *6*(10), e26426.

Wang, D., Xia, Y., Li, X., Hou, L., & Yu, J. (2012). The Rice Genome Knowledgebase (RGKbase): an annotation database for rice comparative genomics and evolutionary biology. *Nucleic Acids Res, 41*(Database issue), D1199-1205. doi: gks1225 [pii], 10.1093/nar/gks1225

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., . . . He, Z. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat Genet, 40*(11), 1370-1374. doi: 10.1038/ng.220

Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., . . . Sasaki, H. (2011). Role for piRNAs and noncoding RNA in *de novo* DNA methylation of the imprinted mouse Rasgrf1 locus. *Science, 332*(6031), 848-852. doi: 10.1126/science.1203919

Waters, A. J., Bilinski, P., Eichten, S. R., Vaughn, M. W., Ross-Ibarra, J., Gehring, M., & Springer, N. M. (2013). Comprehensive analysis of imprinted genes in maize reveals allelic variation for

imprinting and limited conservation with other species. *Proc Natl Acad Sci U S A, 110*(48), 19639-19644. doi: 1309182110 [pii], 10.1073/pnas.1309182110

Waters, A. J., Makarevitch, I., Eichten, S. R., Swanson-Wagner, R. A., Yeh, C. T., Xu, W., . . . Springer, N. M. (2011). Parent-of-Origin Effects on Gene Expression and DNA Methylation in the Maize Endosperm. *Plant Cell, 23*(12), 4221-4233. doi: tpc.111.092668 [pii], 10.1105/tpc.111.092668

Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., & Kohler, C. (2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet, 6*(10). doi: 10.1371/journal.pgen.1001152

Wilkins, J. F., & Haig, D. (2003). What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet, 4*(5), 359-368. doi: 10.1038/nrg1062

Wilson, Z. A., & Zhang, D. B. (2009). From Arabidopsis to rice: pathways in pollen development. *J Exp Bot, 60*(5), 1479-1492. doi: 10.1093/jxb/erp095

Wolf, J. B., & Hager, R. (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol, 4*(12), e380. doi: 10.1371/journal.pbio.0040380

Wolf, J. B., Oakey, R. J., & Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity (Edinb), 113*(2), 167-175. doi: 10.1038/hdy.2014.11

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T., . . . Kohler, C. (2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. *PLoS Genet, 7*(6), e1002126. doi: 10.1371/journal.pgen.1002126

Wu, J. L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M. R., . . . Leung, H. (2005). Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol Biol, 59*(1), 85-97. doi: 10.1007/s11103-004-5112-0

Xiao, W., Gehring, M., Choi, Y., Margossian, L., Pu, H., Harada, J. J., . . . Fischer, R. L. (2003). Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev Cell, 5*(6), 891-901. doi: S1534580703003617 [pii]

Xin, M., Yang, R., Yao, Y., Ma, C., Peng, H., Sun, Q., . . . Ni, Z. (2014). Dynamic parent-of-origin effects on small interfering RNA expression in the developing maize endosperm. *BMC Plant Biol, 14*, 192. doi: 10.1186/s12870-014-0192-8

Xu, W., Dai, M., Li, F., & Liu, A. Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. *Nucleic Acids Res, 42*(11), 6987-6998. doi: 10.1093/nar/gku375

Yadegari, R., & Drews, G. N. (2004). Female gametophyte development. *Plant Cell, 16 Suppl*, S133-141. doi: 10.1105/tpc.018192

Yoshida, T., & Kawabe, A. (2013). Importance of gene duplication in the evolution of genomic imprinting revealed by molecular evolutionary analysis of the type I MADS-box gene family in Arabidopsis species. *PLoS One, 8*(9), e73588. doi: 10.1371/journal.pone.0073588

Zemach, A., Kim, M. Y., Silva, P., Rodrigues, J. A., Dotson, B., Brooks, M. D., & Zilberman, D. (2010). Local DNA hypomethylation activates genes in rice endosperm. *Proc Natl Acad Sci U S A, 107*(43), 18729-18734. doi: 1009695107 [pii], 10.1073/pnas.1009695107

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science, 328*(5980), 916-919. doi: science.1186366 [pii], 10.1126/science.1186366

Zemach, A., & Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol, 20*(17), R780-785. doi: S0960-9822(10)00866-3 [pii], 10.1016/j.cub.2010.07.007

Zhang, H., & Zhu, J. K. (2011). RNA-directed DNA methylation. *Curr Opin Plant Biol, 14*(2), 142-147. doi: S1369-5266(11)00008-2 [pii], 10.1016/j.pbi.2011.02.003

Zhang, H. W., Pan, X. W., Li, Y. C., Wan, L. Y., Li, X. X., & Huang, R. F. (2012). Comparison of differentially expressed genes involved in drought response between two elite rice varieties. *Mol Plant, 5*(6), 1403-1405. doi: 10.1093/mp/sss053

Zhang, M., Xie, S., Dong, X., Zhao, X., Zeng, B., Chen, J., . . . Lai, J. (2014). Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res, 24*(1), 167-176. doi: 10.1101/gr.155879.113

Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., . . . Lai, J. (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A, 108*(50), 20042-20047. doi: 1112186108 [pii], 10.1073/pnas.1112186108

Zhang, W., Wu, Y., Schnable, J. C., Zeng, Z., Freeling, M., Crawford, G. E., & Jiang, J. (2012). High-resolution mapping of open chromatin in the rice genome. *Genome Res, 22*(1), 151-162. doi: gr.131342.111 [pii], 10.1101/gr.131342.111

Zhang, X., & Borevitz, J. O. (2009). Global analysis of allele-specific expression in Arabidopsis thaliana. *Genetics, 182*(4), 943-954. doi: 10.1534/genetics.109.103499

Zhu, J. K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet, 43*, 143-166. doi: 10.1146/annurev-genet-102108-134205

Appendix A: List of imprinted genes identified in rice endosperm

Table A1. 251 paternally expressed genes identified by our pipeline to be imprinted in all assessable varieties. Asterisks indicate those genes also found to be paternally expressed in a study by (Luo et al., 2011).

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os01g02050	phosphoenolpyruvate carboxylase, putative	yes	yes	yes	unknown
LOC_Os01g07920	prolyl 4-hydroxylase, putative	yes	yes	yes	unknown
LOC_Os01g07930	zinc finger C-x8-C-x5-C-x3-H type family protein	yes	yes	yes	unknown
LOC_Os01g08570	2-oxoglutarate and iron-dependent oxygenase domain-containing protein2, putative	yes	yes	yes	unknown
LOC_Os01g11130	RNA recognition motif containing protein, putative	yes	yes	yes	unknown
LOC_Os01g13360	phosphatidylinositol 3- and 4-kinase family protein, putative	yes	yes	yes	unknown
LOC_Os01g13460	helix-loop-helix DNA-binding domain containing protein	yes	yes	yes	yes
LOC_Os01g16810	MYB family transcription factor, putative	yes	yes	yes	unknown
LOC_Os01g17250	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative	yes	yes	yes	unknown
LOC_Os01g17990	expressed protein	yes	yes	yes	unknown
LOC_Os01g27750	bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase,chloroplast precursor, putative	yes	yes	yes	unknown
LOC_Os01g31750	transposon protein, putative, unclassified	unknown	yes	yes	unknown
LOC_Os01g33650	expressed protein	yes	yes	yes	unknown
LOC_Os01g40590	tyrosine protein kinase domain containing protein, putative	yes	yes	yes	unknown
LOC_Os01g41370	FBD domain containing protein, putative	yes	yes	yes	unknown
LOC_Os01g51754	alpha-amylase precursor, putative	yes	unknown	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os01g54100	CK1_CaseinKinase_1a.2 - CK1 includes the casein kinase 1 kinases	yes	yes	yes	unknown
LOC_Os01g54784*	expressed protein	yes	yes	yes	unknown
LOC_Os01g55450	CAMK_KIN1/SNF1/Nim1_like.11 - CAMK includes calcium/calmodulin depedent protein kinases	yes	yes	yes	unknown
LOC_Os01g56100	BSD domain containing protein	yes	yes	yes	unknown
LOC_Os01g56110*	RNA recognition motif containing protein, putative	yes	yes	yes	unknown
LOC_Os01g56580	CK1_CaseinKinase_1a.3 - CK1 includes the casein kinase 1 kinases	yes	yes	yes	unknown
LOC_Os01g57890	Homeobox domain containing protein	yes	yes	unknown	unknown
LOC_Os01g61590	CAMK_CAMK_like.1 - CAMK includes calcium/calmodulin depedent protein kinases	yes	yes	yes	unknown
LOC_Os01g62460	ZOS1-16 - C2H2 zinc finger protein	yes	yes	yes	unknown
LOC_Os01g63250*	josephin, putative	yes	yes	yes	unknown
LOC_Os01g65840	pentatricopeptide, putative	unknown	yes	yes	unknown
LOC_Os01g65850*	CHD3-type chromatin-remodeling factor PICKLE, putative	unknown	yes	yes	unknown
LOC_Os01g67250	Rad21 / Rec8 like protein, putative	yes	yes	yes	unknown
LOC_Os01g67870	expressed protein	yes	yes	yes	unknown
LOC_Os01g68820*	expressed protein	yes	yes	yes	unknown
LOC_Os01g69040*	zinc finger, C3HC4 type domain containing protein	yes	yes	yes	unknown
LOC_Os01g70060*	protein of unknown function, DUF618 domain containing protein	yes	yes	yes	unknown
LOC_Os02g05630	protein phosphatase 2C, putative	yes	yes	yes	unknown
LOC_Os02g07790	serine/threonine-protein kinase HT1, putative	yes	yes	yes	unknown
LOC_Os02g08300	RAD23 DNA repair protein, putative	yes	yes	yes	unknown
LOC_Os02g09650	AP2 domain containing protein	yes	yes	yes	unknown
LOC_Os02g09770	abhydrolase domain-containing protein FAM108C1, putative	yes	yes	yes	unknown
LOC_Os02g10650	CRAL/TRIO domain containing protein	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os02g12870	expressed protein	yes	yes	yes	unknown
LOC_Os02g17190*	MYB family transcription factor, putative	yes	unknown	yes	unknown
LOC_Os02g17910	CK1_CaseinKinase_1.4 - CK1 includes the casein kinase 1 kinases	unknown	yes	yes	unknown
LOC_Os02g21430*	AML1, putative	yes	yes	yes	unknown
LOC_Os02g21460	uncharacterized protein yqjG, putative	yes	yes	yes	unknown
LOC_Os02g34360	expressed protein	yes	yes	yes	unknown
LOC_Os02g34370	expressed protein	yes	yes	yes	unknown
LOC_Os02g34500	expressed protein	yes	yes	yes	unknown
LOC_Os02g34850	histone-lysine N-methyltransferase ASHH2, putative	yes	yes	yes	unknown
LOC_Os02g35010	STE_MEKK_ste11_MAP3K.9 - STE kinases include homologs to sterile 7, sterile 11 and sterile 20 from yeast	yes	yes	yes	unknown
LOC_Os02g36360	ZOS2-11 - C2H2 zinc finger protein	yes	yes	yes	yes
LOC_Os02g36680	expressed protein	yes	yes	yes	unknown
LOC_Os02g39920*	AT hook motif family protein	yes	yes	yes	yes
LOC_Os02g41610	expressed protein	yes	yes	yes	unknown
LOC_Os02g51540*	eukaryotic aspartyl protease domain containing protein	yes	yes	yes	unknown
LOC_Os02g51550	DUF581 domain containing protein	yes	yes	yes	unknown
LOC_Os02g51860*	dehydration response related protein, putative	yes	unknown	yes	unknown
LOC_Os02g54120*	CCR4-NOT transcription factor, putative	yes	unknown	yes	unknown
LOC_Os02g55570	shugoshin-1, putative	unknown	yes	yes	unknown
LOC_Os02g56530	ankyrin repeat domain containing protein	yes	yes	yes	unknown
LOC_Os02g57080*	serine/threonine-protein kinase, putative	yes	yes	yes	unknown
LOC_Os02g57820	AT hook motif domain containing protein	yes	yes	yes	unknown
LOC_Os03g03070	transcription factor, putative	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os03g05480	ZOS3-01 - C2H2 zinc finger protein	yes	yes	yes	unknown
LOC_Os03g07130	RING finger protein 13, putative	yes	yes	yes	unknown
LOC_Os03g14290	expressed protein	yes	yes	yes	unknown
LOC_Os03g14300	THION29 - Plant thionin family protein precursor	yes	yes	yes	unknown
LOC_Os03g15010	transposon protein, putative, unclassified	yes	yes	yes	yes
LOC_Os03g24900	DHHC zinc finger domain containing protein	unknown	yes	yes	unknown
LOC_Os03g25070	CAMK_CAMK_like.18 - CAMK includes calcium/calmodulin depedent protein kinases	yes	yes	yes	unknown
LOC_Os03g27450*	ADP-ribosylation factor, putative	yes	yes	yes	unknown
LOC_Os03g27460*	heat shock protein DnaJ, putative	yes	yes	yes	unknown
LOC_Os03g31070	protein kinase, putative	unknown	yes	yes	unknown
LOC_Os03g36790	tobamovirus multiplication protein, putative	yes	yes	yes	unknown
LOC_Os03g38970*	metal ion binding protein, putative	yes	yes	yes	unknown
LOC_Os03g43580	IQ calmodulin-binding motif family protein, putative	yes	unknown	yes	unknown
LOC_Os03g43590	LSTK-1-like kinase, putative	yes	yes	yes	unknown
LOC_Os03g45194	oxidoreductase, short chain dehydrogenase/reductase family domain containing protein	yes	unknown	yes	unknown
LOC_Os03g45210	2-aminoethanethiol dioxygenase, putative	yes	yes	yes	yes
LOC_Os03g51610	Inositol 1, 3, 4-trisphosphate 5/6-kinase, putative	yes	yes	yes	unknown
LOC_Os03g53630	PHD finger family protein, putative	yes	yes	yes	unknown
LOC_Os03g54900	phytosulfokine receptor precursor, putative	yes	yes	yes	unknown
LOC_Os03g56310	5-nucleotidase domain-containing protein, putative	yes	yes	yes	unknown
LOC_Os03g57560	piwi domain containing protein	unknown	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os03g59680	PAPA-1-like conserved region family protein	yes	yes	yes	unknown
LOC_Os03g59740	ADP-ribosylation factor, putative	yes	yes	unknown	unknown
LOC_Os03g60130	transcription elongation factor protein, putative	yes	yes	yes	unknown
LOC_Os03g60140	U-box domain-containing protein, putative	yes	yes	yes	unknown
LOC_Os03g60150	protein kinase domain containing protein	yes	yes	yes	unknown
LOC_Os03g60710	protein kinase domain containing protein	yes	yes	yes	unknown
LOC_Os03g61120	anthranilate synthase component I-1, chloroplast precursor, putative	yes	yes	yes	unknown
LOC_Os03g63040	expressed protein	yes	yes	yes	unknown
LOC_Os03g63770	RCD1, putative	yes	yes	yes	unknown
LOC_Os04g05030	serine-rich 25 kDa antigen protein, putative	yes	yes	yes	unknown
LOC_Os04g08470	OsFBX116 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os04g10214	expressed protein	yes	unknown	yes	unknown
LOC_Os04g11830	TCP family transcription factor, putative	yes	unknown	yes	unknown
LOC_Os04g19080	retrotransposon protein, putative, Ty3-gypsy subclass	yes	unknown	yes	unknown
LOC_Os04g20774*	pumilio-family RNA binding repeat containing protein	yes	yes	yes	unknown
LOC_Os04g20800	pumilio domain-containing protein, putativ	yes	yes	unknown	unknown
LOC_Os04g22240*	zinc finger, C3HC4 type domain containing protein	yes	yes	yes	unknown
LOC_Os04g31170	expressed protein	yes	yes	unknown	unknown
LOC_Os04g32250	expressed protein	unknown	yes	yes	unknown
LOC_Os04g32880*	CBS domain containing membrane protein, putative	unknown	yes	yes	unknown
LOC_Os04g32970	OTU-like cysteine protease family protein, putative	yes	yes	yes	unknown
LOC_Os04g40660	MA3 domain containing protein	yes	yes	yes	yes

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os04g41470	expressed protein	yes	yes	yes	unknown
LOC_Os04g42250*	transferase family protein, putative	yes	yes	yes	unknown
LOC_Os04g42260	protein phosphatase 2C, putative	yes	unknown	yes	unknown
LOC_Os04g54420	protein of unknown function, DUF618 domain containing protein	yes	yes	yes	yes
LOC_Os04g56720	RCC2, putative	yes	yes	yes	unknown
LOC_Os04g56800	peptidyl-prolyl cis-trans isomerase, putative	unknown	yes	yes	unknown
LOC_Os04g57640	RCD1, putative	yes	yes	yes	unknown
LOC_Os05g01210	expressed protein	yes	yes	yes	unknown
LOC_Os05g01230	zinc finger, C3HC4 type domain containing protein	yes	yes	yes	unknown
LOC_Os05g01240	AML1, putative	yes	yes	yes	unknown
LOC_Os05g01710	transcription initiation factor IIA gamma chain, putative	unknown	yes	yes	unknown
LOC_Os05g03630	dnaJ domain containing protein	yes	yes	yes	unknown
LOC_Os05g04330	DNA methyltransferase protein, putative	yes	yes	yes	unknown
LOC_Os05g04520	protein kinase, putative	yes	yes	yes	unknown
LOC_Os05g05780*	chromatin-remodeling complex ATPase chain, putative	yes	yes	yes	unknown
LOC_Os05g05790*	double-stranded RNA binding motif containing protein	yes	yes	yes	unknown
LOC_Os05g06260	Spc97 / Spc98 family protein, putative	yes	yes	yes	unknown
LOC_Os05g28180	AMP deaminase, putative	yes	yes	yes	unknown
LOC_Os05g31380	GRAS family transcription factor containing protein	yes	yes	yes	unknown
LOC_Os05g34510	expressed protein	yes	yes	yes	unknown
LOC_Os05g41220*	SNF1-related protein kinase regulatory subunit beta-1, putative	yes	yes	yes	unknown
LOC_Os05g43480	ubiquitin carboxyl-terminal hydrolase domain containing protein	yes	yes	yes	unknown
LOC_Os05g45060	RING-H2 finger protein ATL2M, putative	yes	unknown	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os05g48560*	expressed protein	yes	yes	yes	unknown
LOC_Os05g48980	ras-related protein, putative	yes	yes	yes	unknown
LOC_Os05g50970	protein phosphatase 2C, putative	yes	yes	yes	unknown
LOC_Os05g51400	protein kinase APK1B, chloroplast precursor, putative	yes	yes	yes	unknown
LOC_Os06g01680	expressed protein	yes	yes	yes	unknown
LOC_Os06g02028	eyes absent homolog 4, putative	yes	yes	yes	unknown
LOC_Os06g03860	uncharacterized membrane protein, putative	yes	yes	yes	unknown
LOC_Os06g04920	zinc finger family protein, putative	yes	unknown	yes	unknown
LOC_Os06g04930	expressed protein	yes	yes	yes	unknown
LOC_Os06g06870*	zinc finger protein, putative	yes	yes	yes	unknown
LOC_Os06g08740*	expressed protein	yes	yes	yes	unknown
LOC_Os06g11620	RNA recognition motif containing protein, putative	yes	yes	yes	unknown
LOC_Os06g12590	protein kinase, putative	yes	unknown	yes	unknown
LOC_Os06g12680	RING-H2 finger protein, putative	yes	yes	yes	unknown
LOC_Os06g19660	WD domain, G-beta repeat domain containing protein	yes	unknown	yes	unknown
LOC_Os06g21900	expressed protein	yes	yes	yes	unknown
LOC_Os06g30060	expressed protein	yes	yes	yes	unknown
LOC_Os06g35530	CGMC_GSK.8 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases	yes	yes	yes	yes
LOC_Os06g37670	S-locus-like receptor protein kinase, putative	unknown	unknown	yes	yes
LOC_Os06g39760	WD domain, G-beta repeat domain containing protein	yes	yes	yes	yes
LOC_Os06g40490*	glycosyl hydrolases family 17, putative	yes	yes	yes	unknown
LOC_Os06g42990*	AP2 domain containing protein	yes	yes	yes	unknown
LOC_Os06g44034	expressed protein	unknown	yes	yes	unknown
LOC_Os06g45000	ubiquitin-conjugating enzyme, putative	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os06g47290	growth regulator related protein, putative	yes	yes	yes	unknown
LOC_Os06g47294	methyltransferase, putative	yes	unknown	yes	unknown
LOC_Os07g06980	histone deacetylase, putative	yes	yes	yes	unknown
LOC_Os07g12490*	KH domain containing protein, putative	yes	yes	yes	unknown
LOC_Os07g12630	transcription elongation factor protein, putative	yes	yes	yes	unknown
LOC_Os07g17460*	OsFBL36 - F-box domain and LRR containing protein	yes	yes	yes	unknown
LOC_Os07g18710	OsFBLD8 - F-box, LRR and FBD domain containing protein	yes	unknown	yes	unknown
LOC_Os07g18720	tetratricopeptide repeat containing protein, putative	yes	yes	yes	unknown
LOC_Os07g27110	RNA recognition motif containing protein	yes	yes	yes	unknown
LOC_Os07g27300*	RNA-binding protein Luc7-like, putative	yes	yes	yes	unknown
LOC_Os07g32412	transposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os07g38080	homeodomain-like, putative	yes	yes	yes	unknown
LOC_Os07g41160*	protein of unknown function DUF1675 domain containing protein	yes	yes	yes	unknown
LOC_Os07g42104	expressed protein	unknown	yes	yes	unknown
LOC_Os07g42760	expressed protein	yes	yes	yes	unknown
LOC_Os07g44840	bacterial transferase hexapeptide domain containing protein	yes	yes	yes	unknown
LOC_Os07g48170	nucleotidyltransferase, putative	yes	unknown	yes	unknown
LOC_Os07g48200	B3 DNA binding domain containing protein, putative	yes	yes	yes	unknown
LOC_Os07g48229	vacuolar-sorting receptor precursor, putative	unknown	yes	yes	unknown
LOC_Os07g48260	WRKY47	unknown	yes	yes	unknown
LOC_Os08g01040	zinc finger, C3HC4 type domain containing protein	yes	yes	yes	unknown
LOC_Os08g01054*	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os08g02690	MA3 domain containing protein	yes	yes	yes	unknown
LOC_Os08g08000	DNA binding protein, putative	yes	yes	yes	unknown
LOC_Os08g15030	aspartate carbamoyltransferase, putative	unknown	yes	yes	unknown
LOC_Os08g24400	SWP, putative	yes	yes	yes	yes
LOC_Os08g24420*	SWP, putative	yes	yes	yes	yes
LOC_Os08g24930	expressed protein	unknown	yes	yes	unknown
LOC_Os08g25640	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os08g27240*	ARID/BRIGHT DNA-binding domain containing protein	yes	yes	yes	yes
LOC_Os08g35050	ARID/BRIGHT DNA-binding domain containing protein	yes	yes	yes	unknown
LOC_Os08g40620	rabGAP/TBC domain-containing protein, putative	yes	yes	yes	unknown
LOC_Os08g41030	AP2 domain containing protein	yes	yes	yes	unknown
LOC_Os08g41700	expressed protein	yes	yes	yes	unknown
LOC_Os08g41710*	FHA domain containing protein, putative	yes	yes	yes	unknown
LOC_Os08g41790	phosphoribulokinase/Uridine kinase family protein	yes	yes	yes	unknown
LOC_Os08g42980	la domain containing protein	yes	yes	yes	unknown
LOC_Os08g44020	rhamnogalacturonate lyase, putative	unknown	yes	yes	unknown
LOC_Os08g45130	histone-lysine N-methyltransferase, putative	yes	yes	yes	unknown
LOC_Os09g03090*	expressed protein	yes	yes	yes	unknown
LOC_Os09g11760	expressed protein	yes	yes	yes	unknown
LOC_Os09g14680	expressed protein	unknown	yes	yes	unknown
LOC_Os09g20010	SNF1-related protein kinase regulatory subunit beta-2, putative	yes	yes	yes	unknown
LOC_Os09g20650*	OsFBX323 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os09g23720	expressed protein	unknown	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os09g24954	double-stranded RNA binding motif containing protein	yes	unknown	yes	unknown
LOC_Os09g28940*	ubiquitin carboxyl-terminal hydrolase domain containing protein	yes	unknown	yes	unknown
LOC_Os10g01480	Inositol 1, 3, 4-trisphosphate 5/6-kinase, putative	yes	yes	yes	unknown
LOC_Os10g04610	OsFBX360 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os10g04890*	expressed protein	yes	yes	yes	yes
LOC_Os10g04900	OsFBX364 - F-box domain containing protein	yes	yes	yes	yes
LOC_Os10g04980*	OsFBX365 - F-box domain containing protein	yes	yes	yes	yes
LOC_Os10g05500	OsFBX370 - F-box domain containing protein	yes	yes	yes	yes
LOC_Os10g05530	OsFBX372 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os10g11260	ubiquitin-conjugating enzyme, putative	yes	yes	yes	unknown
LOC_Os10g21196	chloroplast 30S ribosomal protein S16, putative	unknown	yes	yes	unknown
LOC_Os10g26520	protein kinase domain containing protein	yes	unknown	yes	unknown
LOC_Os10g29549	expressed protein	yes	yes	yes	unknown
LOC_Os10g36710	CAMK_CAMK_like.40 - CAMK includes calcium/calmodulin depedent protein kinases	yes	yes	yes	unknown
LOC_Os10g37540	OsFBDUF48 - F-box and DUF domain containing protein	yes	unknown	yes	unknown
LOC_Os10g39780*	protein phosphatase 2C, putative	yes	yes	yes	unknown
LOC_Os10g41360	ARABIDILLO-1, putative	yes	yes	yes	unknown
LOC_Os10g41370	WD repeat-containing protein 8, putative	yes	yes	yes	unknown
LOC_Os10g42690	jmjC domain containing protein	yes	yes	yes	unknown
LOC_Os11g05010	heavy-metal-associated domain-containing protein, putative	yes	unknown	yes	unknown
LOC_Os11g07910*	transmembrane 9 superfamily member, putative	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os11g13430	RGH1A, putative	yes	unknown	yes	unknown
LOC_Os11g13694	AGAP007117-PA, putative	yes	unknown	yes	unknown
LOC_Os11g14300	la domain containing protein	yes	yes	yes	unknown
LOC_Os11g31330	no apical meristem protein, putative	yes	yes	yes	unknown
LOC_Os11g31340	no apical meristem protein, putative	yes	yes	yes	unknown
LOC_Os11g31360	no apical meristem protein, putative	yes	yes	yes	yes
LOC_Os11g31380	no apical meristem protein, putative	yes	yes	yes	unknown
LOC_Os11g34190	expressed protein	yes	yes	yes	unknown
LOC_Os11g36470	ubiquitin carboxyl-terminal hydrolase 21, putative	unknown	yes	yes	unknown
LOC_Os11g37520	BTBT3 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain with tetratricopeptide repeats	yes	yes	yes	yes
LOC_Os11g38630	expressed protein	unknown	unknown	yes	yes
LOC_Os11g38900	histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH1, putative	yes	yes	yes	unknown
LOC_Os11g38980	F-box/Kelch-repeat protein, putative	yes	yes	yes	yes
LOC_Os11g45590	transposon protein, putative, CACTA, En/Spm sub-class	yes	yes	yes	yes
LOC_Os11g45950	NAC domain-containing protein 90, putative	unknown	unknown	yes	yes
LOC_Os12g05040*	heavy-metal-associated domain-containing protein, putative	yes	yes	yes	yes
LOC_Os12g06630	OsFBT14 - F-box and tubby domain containing protein	yes	yes	yes	unknown
LOC_Os12g06640	homeodomain, putative	yes	yes	yes	unknown
LOC_Os12g07120	GATA zinc finger domain containing protein	yes	yes	yes	unknown
LOC_Os12g08720	H-BTB8 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain with H family conserved sequence	yes	yes	yes	unknown
LOC_Os12g08780*	flavin monooxygenase, putative	yes	yes	yes	unknown
LOC_Os12g31350*	SSXT protein, putative	yes	yes	yes	yes

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os12g32150*	expressed protein	yes	yes	yes	unknown
LOC_Os12g32170*	expressed protein	yes	yes	yes	unknown
LOC_Os12g32180	cornichon protein, putative	yes	yes	yes	unknown
LOC_Os12g36810	polygalacturonase, putative	yes	unknown	yes	unknown
LOC_Os12g37860*	expressed protein	yes	yes	yes	unknown
LOC_Os12g38590	pumilio-family RNA binding repeat containing protein	yes	yes	yes	unknown
LOC_Os12g39420	nucleobase-ascorbate transporter, putative	yes	yes	yes	unknown
LOC_Os12g39640	MYB family transcription factor, putative	yes	yes	yes	unknown
LOC_Os12g40510	calcineurin B, putative	yes	yes	yes	unknown
LOC_Os12g40520*	MATH domain containing protein	yes	yes	yes	unknown
LOC_Os12g42310	serine/threonine-protein phosphatase BSL2, putative	yes	yes	yes	unknown

Table A2. 21 additional paternally expressed genes identified by manual inspection of the list of 174 candidate genes for imprinting variation. Asterisks indicate those genes also found to be paternally expressed in a study by (Luo et al., 2011).

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	IR64	Nip	Kit
LOC_Os01g07740*	DEAD-box ATP-dependent RNA helicase 14, putative	yes	yes	yes	unknown
LOC_Os01g16110	la domain containing protein, putative	yes	yes	yes	yes
LOC_Os01g34890	GLUCAN SYNTHASE-LIKE protein, putative	yes	yes	yes	unknown
LOC_Os01g62780	expressed protein	yes	yes	yes	unknown
LOC_Os01g65986	DUF803 domain containing, putative	yes	yes	yes	unknown
LOC_Os01g69850	OsMADS65 - MADS-box family gene with MIKC* type-box	yes	yes	yes	yes
LOC_Os01g73460	ATXR, putative	yes	yes	yes	unknown
LOC_Os02g19220	expressed protein	yes	yes	yes	yes
LOC_Os02g36670	expressed protein	yes	yes	yes	unknown
LOC_Os02g57190*	Myosin head domain containing protein	yes	yes	yes	unknown
LOC_Os03g40930	expressed protein	yes	yes	yes	unknown
LOC_Os03g58480	seed specific protein Bn15D14A, putative	yes	yes	yes	unknown
LOC_Os04g06770	argonaute, putative	yes	yes	yes	unknown
LOC_Os04g42600	polyadenylate-binding protein, putative	yes	yes	yes	unknown
LOC_Os05g02260	expressed protein	yes	yes	yes	unknown
LOC_Os05g32570	RING E3 ligase protein, putative	yes	yes	yes	yes
LOC_Os05g34540	rab GDP dissociation inhibitor alpha, putative	yes	yes	yes	unknown
LOC_Os08g13350	expressed protein	yes	yes	yes	unknown
LOC_Os08g43380	TBC domain containing protein	yes	yes	yes	unknown
LOC_Os08g44860*	aminopeptidase, putative	yes	yes	yes	unknown
LOC_Os11g06580	OsFBX397 - F-box domain containing protein	yes	yes	yes	unknown

Table A3. 189 maternally expressed genes identified by our pipeline to be imprinted in all assessable varieties. Asterisks indicate those genes also found to be maternally expressed in a study by (Luo et al., 2011).

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os01g01070	expressed protein	unknown	unknown	yes	yes
LOC_Os01g01470	no apical meristem protein, putative	yes	yes	yes	unknown
LOC_Os01g05590	retrotransposon protein, putative, unclassified	unknown	unknown	yes	yes
LOC_Os01g08020	boron transporter protein, putative	unknown	yes	yes	unknown
LOC_Os01g10080*	expressed protein	yes	yes	yes	unknown
LOC_Os01g10504	OsMADS3 - MADS-box family gene with MIKCc type-box	yes	unknown	yes	unknown
LOC_Os01g10780	expressed protein	yes	unknown	yes	unknown
LOC_Os01g12890*	expressed protein	yes	yes	yes	unknown
LOC_Os01g13180*	expressed protein	yes	yes	yes	unknown
LOC_Os01g28300	OsFBX7 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os01g28810	expressed protein	unknown	unknown	yes	yes
LOC_Os01g33324	retrotransposon protein, putative, Ty1-copia subclass	unknown	yes	yes	unknown
LOC_Os01g34400	transposon protein, putative, Pong sub-class	yes	yes	yes	unknown
LOC_Os01g35860	transposon protein, putative, Mutator sub-class	yes	unknown	yes	unknown
LOC_Os01g38650*	expressed protein	yes	yes	yes	unknown
LOC_Os01g40450*	2-aminoethanethiol dioxygenase, putative	yes	yes	yes	unknown
LOC_Os01g42090	nodulin MtN3 family protein, putative	yes	yes	yes	unknown
LOC_Os01g42210	LTPL47 - Protease inhibitor/seed storage/LTP family protein precursor, putative	yes	yes	yes	unknown
LOC_Os01g46910	microneme protein Sm70, putative	yes	yes	yes	unknown
LOC_Os01g47470	serine/threonine-protein kinase, putative	yes	yes	yes	unknown
LOC_Os01g48580	ubiquitin-conjugating enzyme E2, putative	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os01g51260	MYB family transcription factor, putative	unknown	yes	yes	unknown
LOC_Os01g52130	sulfate transporter, putative	yes	unknown	yes	unknown
LOC_Os01g52240	chlorophyll A-B binding protein, putative	yes	unknown	yes	unknown
LOC_Os01g56800	MATH domain containing protein	yes	yes	yes	unknown
LOC_Os01g60360	ubiquitin-conjugating enzyme, putative	unknown	yes	yes	unknown
LOC_Os01g64810*	zinc finger DHHC domain-containing protein, putative	yes	yes	yes	unknown
LOC_Os01g71280	glycerol-3-phosphate dehydrogenase, putative	yes	yes	yes	unknown
LOC_Os01g72610	glycosyltransferase, putative	yes	unknown	yes	unknown
LOC_Os01g74140	WRKY17	yes	unknown	yes	unknown
LOC_Os02g01840	retrotransposon protein, putative, unclassified	yes	unknown	yes	unknown
LOC_Os02g10890	expressed protein	yes	unknown	yes	unknown
LOC_Os02g14720	expressed protein	yes	yes	yes	unknown
LOC_Os02g16544	expressed protein	yes	yes	yes	yes
LOC_Os02g19130	transcription factor X1, putative	yes	unknown	yes	unknown
LOC_Os02g26380	expressed protein	yes	yes	yes	unknown
LOC_Os02g27720	retrotransposon protein, putative, unclassified	unknown	unknown	yes	yes
LOC_Os02g28580*	expressed protein	yes	yes	yes	yes
LOC_Os02g28660*	expressed protein	yes	yes	yes	yes
LOC_Os02g32210	expressed protein	unknown	yes	yes	unknown
LOC_Os02g34990	ACT domain containing protein, putative	yes	yes	yes	unknown
LOC_Os02g38120	BTBN3 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain with non-phototropic hypocotyl 3 NPH3 domain	yes	yes	yes	unknown
LOC_Os02g41670	phenylalanine ammonia-lyase, putative	unknown	yes	yes	unknown
LOC_Os02g41810	expressed protein	yes	yes	yes	yes
LOC_Os02g43460*	required to maintain repression 1, putative	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os02g45770	OsMADS6 - MADS-box family gene with MIKCc type-box	yes	yes	yes	unknown
LOC_Os02g46680	multidrug resistance protein, putative	yes	unknown	yes	unknown
LOC_Os02g48820	plastocyanin-like domain containing protein, putative	yes	yes	yes	unknown
LOC_Os02g53730	expressed protein	unknown	yes	yes	unknown
LOC_Os02g55560*	protein phosphatase 2C, putative	yes	yes	yes	unknown
LOC_Os03g01320*	LTPL116 - Protease inhibitor/seed storage/LTP family protein precursor	yes	yes	yes	unknown
LOC_Os03g04650	cytochrome P450 protein, putative	yes	unknown	yes	unknown
LOC_Os03g15320	glyoxal oxidase-related, putative	unknown	unknown	yes	yes
LOC_Os03g16740	protein kinase APK1A, chloroplast precursor, putative	yes	unknown	yes	unknown
LOC_Os03g16760	protein phosphatase 2C, putative	yes	yes	yes	unknown
LOC_Os03g21870	DUF623 domain containing protein	yes	yes	yes	unknown
LOC_Os03g22470	desiccation-related protein PCC13-62 precursor, putative	yes	yes	yes	unknown
LOC_Os03g31944	expressed protein	yes	yes	yes	unknown
LOC_Os03g35570	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os03g38170	expressed protein	yes	unknown	yes	unknown
LOC_Os03g58530	ES43 protein, putative	yes	yes	yes	unknown
LOC_Os03g58790	ATPase, putative	yes	unknown	yes	unknown
LOC_Os04g07740	retrotransposon protein, putative, unclassified	unknown	unknown	yes	yes
LOC_Os04g07830	expressed protein	unknown	yes	yes	unknown
LOC_Os04g08570*	uncharacterized PE-PGRS family protein PE_PGRS20, putative	yes	yes	yes	unknown
LOC_Os04g10260	basic region leucine zipper domain containing protein	yes	yes	yes	unknown
LOC_Os04g18030	Sec1 family transport protein, putative	yes	unknown	yes	unknown
LOC_Os04g19870	expressed protein	unknown	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os04g28030	expressed protein	unknown	unknown	yes	yes
LOC_Os04g28120*	response regulator receiver domain containing protein	unknown	yes	yes	unknown
LOC_Os04g28210	verticillium wilt disease resistance protein, putative	unknown	yes	yes	unknown
LOC_Os04g38525	expressed protein	yes	yes	yes	unknown
LOC_Os04g39150*	pathogenesis-related Bet v I family protein, putative	yes	yes	yes	unknown
LOC_Os04g39560*	expressed protein	yes	yes	yes	unknown
LOC_Os04g56700	naringenin,2-oxoglutarate 3-dioxygenase, putative	yes	yes	yes	unknown
LOC_Os05g03640	flavonol synthase/flavanone 3-hydroxylase, putative	yes	yes	yes	unknown
LOC_Os05g05660	PWWP domain containing protein	yes	yes	yes	unknown
LOC_Os05g10800*	expressed protein	yes	yes	yes	unknown
LOC_Os05g23950	TRAF-type zinc finger family protein	yes	yes	yes	unknown
LOC_Os05g26040*	pumilio-family RNA binding repeat containing protein	yes	yes	yes	unknown
LOC_Os05g26110*	transposon protein, putative, Mutator sub-class	yes	yes	yes	unknown
LOC_Os05g27730	WRKY53	yes	unknown	yes	unknown
LOC_Os05g34310*	no apical meristem protein, putative	yes	yes	yes	unknown
LOC_Os05g40790*	CCR4-NOT transcription factor, putative	yes	yes	yes	unknown
LOC_Os05g48280	expressed protein	yes	yes	yes	unknown
LOC_Os05g48610	expressed protein	yes	yes	yes	unknown
LOC_Os05g50390*	expressed protein	yes	yes	yes	unknown
LOC_Os06g04169	hydrolase, alpha/beta fold family domain containing protein	yes	unknown	yes	unknown
LOC_Os06g05550	GDSL-like lipase/acylhydrolase, putative	yes	yes	yes	unknown
LOC_Os06g11730*	RNA recognition motif containing protein, putative	yes	yes	yes	unknown
LOC_Os06g30280	expressed protein	yes	yes	yes	unknown
LOC_Os06g33640*	CAPIP1, putative	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os06g33690*	CAPIP1, putative	yes	yes	yes	unknown
LOC_Os06g40020	DEAD-box ATP-dependent RNA helicase 52A, putative	yes	yes	yes	unknown
LOC_Os06g40150	AP2 domain containing protein	yes	yes	yes	unknown
LOC_Os06g45640	core histone H2A/H2B/H3/H4, putative	yes	yes	yes	unknown
LOC_Os07g08880	ES43 protein, putative	yes	yes	yes	unknown
LOC_Os07g09020*	argonaute, putative	yes	yes	yes	unknown
LOC_Os07g12130	MYB family transcription factor, putative	yes	yes	yes	unknown
LOC_Os07g12260*	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os07g20110	expressed protein	yes	yes	yes	unknown
LOC_Os07g20120	expressed protein	unknown	yes	yes	unknown
LOC_Os07g27030	OsFBX236 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os07g32360*	expressed protein	yes	yes	yes	unknown
LOC_Os07g32370	transposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os07g33710	expressed protein	yes	unknown	yes	unknown
LOC_Os07g34620*	expressed protein	yes	yes	yes	yes
LOC_Os07g36920	OsFBX255 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os07g39970	ZOS7-08 - C2H2 zinc finger protein	yes	unknown	yes	unknown
LOC_Os07g41650*	pectinesterase, putative	yes	unknown	yes	unknown
LOC_Os07g47160	OsFBX259 - F-box domain containing protein	yes	yes	yes	yes
LOC_Os07g47430	expressed protein	unknown	unknown	yes	yes
LOC_Os07g48030	peroxidase precursor, putative	yes	yes	yes	unknown
LOC_Os07g48400	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os08g01570	retrotransposon protein, putative, unclassified	unknown	yes	yes	yes

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os08g03470*	MBTB15 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain with Meprin and TRAF Homology MATH domain	yes	yes	yes	unknown
LOC_Os08g04290*	WD domain, G-beta repeat domain containing protein	yes	yes	yes	yes
LOC_Os08g04470	U box protein 8, putative	yes	yes	yes	yes
LOC_Os08g07710	hypothetical protein	yes	yes	yes	yes
LOC_Os08g08960*	Cupin domain containing protein	yes	yes	yes	unknown
LOC_Os08g09700*	OsFBX270 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os08g15840	ankyrin repeat-rich protein, putative	yes	yes	yes	unknown
LOC_Os08g19590	Homeobox domain containing protein	yes	yes	yes	unknown
LOC_Os08g27870*	EARLY flowering protein, putative	yes	yes	yes	unknown
LOC_Os08g28710	receptor protein kinase CRINKLY4 precursor, putative	yes	unknown	yes	unknown
LOC_Os08g28940	OsFBX289 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os08g28960*	expressed protein	yes	yes	yes	unknown
LOC_Os08g29580	expressed protein	yes	yes	yes	unknown
LOC_Os08g31340*	heavy metal-associated domain containing protein	yes	unknown	yes	unknown
LOC_Os08g38850*	phosphatidylinositol transfer, putative	yes	yes	yes	unknown
LOC_Os08g42350*	nodulin MtN3 family protein, putative	yes	yes	yes	unknown
LOC_Os08g44180	expressed protein	unknown	yes	yes	unknown
LOC_Os09g02690*	expressed protein	yes	yes	yes	unknown
LOC_Os09g03500*	ZOS9-01 - C2H2 zinc finger protein	yes	yes	yes	unknown
LOC_Os09g06950	expressed protein	yes	yes	yes	unknown
LOC_Os09g07940*	retrotransposon protein, putative, unclassified	yes	unknown	yes	unknown
LOC_Os09g08300	expressed protein	unknown	yes	yes	unknown
LOC_Os09g17800	retrotransposon protein, putative, unclassified	yes	unknown	yes	unknown
LOC_Os09g25160	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os09g25170	retrotransposon protein, putative, unclassified	unknown	yes	yes	yes
LOC_Os09g32948	OsMADS8 - MADS-box family gene with MIKCc type-box	yes	yes	yes	unknown
LOC_Os09g32992	expressed protein	unknown	yes	yes	unknown
LOC_Os09g34880*	basic region leucine zipper domain containing protein	yes	yes	yes	unknown
LOC_Os09g36470*	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os09g36970	expressed protein	yes	yes	yes	unknown
LOC_Os09g38429*	hypothetical protein	unknown	yes	yes	unknown
LOC_Os10g03870	OsFBX353 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os10g05750*	POEI3 - Pollen Ole e I allergen and extensin family protein precursor	yes	yes	yes	unknown
LOC_Os10g10760	expressed protein	yes	unknown	yes	unknown
LOC_Os10g13460	hypothetical protein	yes	yes	yes	unknown
LOC_Os10g26430	agenet domain containing protein	yes	yes	yes	unknown
LOC_Os10g29560	transposon protein, putative, CACTA, En/Spm sub-class	yes	yes	yes	unknown
LOC_Os10g32580	GDSL-like lipase/acylhydrolase, putative	yes	yes	yes	unknown
LOC_Os10g37830*	OsFBX391 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os10g39420*	CAMK_CAMK_like.8 - CAMK includes calcium/calmodulin depedent protein kinases	yes	yes	yes	unknown
LOC_Os11g03220	RNA binding protein, putative	yes	yes	yes	unknown
LOC_Os11g06780	serine/threonine-protein kinase BRI1-like 1 precursor, putative	yes	unknown	yes	unknown
LOC_Os11g10460	peroxidase precursor, putative	yes	unknown	yes	unknown
LOC_Os11g11200	expressed protein	unknown	unknown	yes	yes
LOC_Os11g14150	transposon protein, putative, Pong sub-class	yes	unknown	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os11g15620	OsFBX420 - F-box domain containing protein	yes	yes	yes	unknown
LOC_Os11g25470	expressed protein	yes	yes	yes	unknown
LOC_Os11g26030	GTPase of unknown function domain containing protein, putative	unknown	yes	yes	unknown
LOC_Os11g26830*	ATP-binding region, ATPase-like domain containing protein	unknown	yes	yes	unknown
LOC_Os11g27470*	expressed protein	yes	yes	yes	unknown
LOC_Os11g35400	AMP-binding enzyme, putative	yes	yes	yes	unknown
LOC_Os11g36760*	transmembrane receptor, putative	yes	yes	yes	unknown
LOC_Os11g36770	expressed protein	yes	unknown	yes	unknown
LOC_Os11g37730	glutathione S-transferase, N-terminal domain containing protein	unknown	yes	yes	unknown
LOC_Os11g40530	LTPL162 - Protease inhibitor/seed storage/LTP family protein precursor	yes	yes	yes	unknown
LOC_Os11g42890	expressed protein	unknown	unknown	yes	yes
LOC_Os12g06480	PHD-finger family protein	yes	yes	yes	unknown
LOC_Os12g06950	hypothetical protein	yes	yes	yes	unknown
LOC_Os12g10540	OsMADS13 - MADS-box family gene with MIKCc type-box	yes	yes	yes	unknown
LOC_Os12g10800	transposon protein, putative, unclassified	yes	yes	yes	yes
LOC_Os12g13100	WW domain containing protein	yes	unknown	yes	unknown
LOC_Os12g13450	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os12g13670	expressed protein	yes	yes	yes	unknown
LOC_Os12g16600	expressed protein	yes	unknown	yes	unknown
LOC_Os12g17430	NBS-LRR disease resistance protein, putative	yes	yes	yes	unknown
LOC_Os12g17670	expressed protein	yes	yes	yes	unknown
LOC_Os12g17710	expressed protein	yes	yes	yes	unknown
LOC_Os12g18960	integral membrane protein DUF6 containing protein	yes	unknown	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os12g22030	serine hydroxymethyltransferase, mitochondrial precursor, putative	yes	yes	yes	unknown
LOC_Os12g22680	histone H3, putative	yes	yes	yes	unknown
LOC_Os12g23040	transposon protein, putative, Mutator sub-class	yes	yes	yes	unknown
LOC_Os12g26420	retrotransposon protein, putative, unclassified	yes	unknown	yes	unknown
LOC_Os12g27994*	expressed protein	yes	yes	yes	unknown
LOC_Os12g39890	expressed protein	unknown	yes	yes	unknown
LOC_Os12g41140	expressed protein	yes	yes	yes	unknown
LOC_Os12g42610	YABBY domain containing protein, putative	yes	yes	yes	unknown

Table A4. 53 additional maternally expressed genes identified by manual inspection of the list of 174 candidate genes for imprinting variation. Asterisks indicate those genes also found to be maternally expressed in a study by (Luo et al., 2011).

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os01g10520*	expressed protein	yes	yes	yes	unknown
LOC_Os01g34410	transposon protein, putative, Pong sub-class	yes	yes	yes	unknown
LOC_Os01g42960	TTL1, putative	yes	yes	yes	unknown
LOC_Os01g43844	cytochrome P450 72A1, putative	yes	yes	yes	unknown
LOC_Os01g45000	expressed protein	yes	yes	yes	unknown
LOC_Os01g49330	formyl transferase, putative	yes	yes	yes	unknown
LOC_Os01g56850	expressed protein	yes	yes	yes	unknown
LOC_Os01g65550	RNA recognition motif containing protein, putative	yes	yes	yes	unknown
LOC_Os01g65770	expressed protein	yes	yes	yes	unknown
LOC_Os02g09500	expressed protein	yes	yes	yes	unknown
LOC_Os02g29230*	expressed protein	yes	yes	yes	unknown
LOC_Os02g31140*	major ampullate spidroin 2-2, putative	yes	yes	yes	yes
LOC_Os02g31970*	retrotransposon protein, putative, unclassified	yes	yes	yes	yes
LOC_Os02g40000	trafficking protein particle complex subunit, putative	yes	yes	yes	unknown
LOC_Os02g43090	myristoyl-acyl carrier protein thioesterase, chloroplast precursor, putative	yes	yes	yes	unknown
LOC_Os03g01490	expressed protein	yes	yes	yes	unknown
LOC_Os03g12350	two-component response regulator, putative	yes	yes	yes	unknown
LOC_Os03g20430*	histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH8, putative	yes	yes	yes	unknown
LOC_Os03g22900*	SNF2 family N-terminal domain containing protein	yes	yes	yes	unknown
LOC_Os03g26020	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os03g30740	expressed protein	yes	yes	yes	unknown
LOC_Os03g43720	transporter family protein, putative	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os03g52700	expressed protein	yes	yes	yes	unknown
LOC_Os04g08034	ZOS4-02 - C2H2 zinc finger protein	yes	yes	yes	yes
LOC_Os04g35000	expressed protein	yes	yes	yes	unknown
LOC_Os05g30150	amino acid transporter family protein, putative	yes	yes	yes	unknown
LOC_Os05g34240	hAT dimerisation domain containing protein	yes	yes	yes	unknown
LOC_Os05g34670	KH domain containing protein, putative	yes	yes	yes	unknown
LOC_Os05g38260	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os05g38600	ZOS5-09 - C2H2 zinc finger protein	yes	yes	yes	unknown
LOC_Os05g47820*	expressed protein	yes	yes	yes	unknown
LOC_Os06g15430	expressed protein	yes	yes	yes	unknown
LOC_Os06g33330	powdery mildew resistant protein 5, putative	yes	yes	yes	yes
LOC_Os06g35970*	meiosis 5, putative	yes	yes	yes	unknown
LOC_Os06g43620	haemolysin-III, putative	yes	yes	yes	unknown
LOC_Os06g48320	EDR2, putative	yes	yes	yes	unknown
LOC_Os07g11010	protein Kinase-associated protein phosphatase, putative	yes	yes	yes	unknown
LOC_Os07g13834	expressed protein	yes	yes	yes	unknown
LOC_Os07g25810	retrotransposon protein, putative, unclassified	yes	yes	yes	unknown
LOC_Os08g17370	transmembrane 9 superfamily member, putative	yes	yes	yes	unknown
LOC_Os09g17680	NAD kinase, putative	yes	yes	yes	unknown
LOC_Os09g32210	pumilio-family RNA binding repeat containing protein	yes	yes	yes	yes
LOC_Os10g01380	CW7, putative	yes	yes	yes	unknown
LOC_Os10g11580	histone-like transcription factor and archaeal histone, putative	yes	yes	yes	unknown
LOC_Os10g27370	transposon protein, putative, Pong sub-class	yes	yes	yes	unknown

Gene	Putative or annotated function	Imprinted in cultivar?			
		IR64	93-11	Nip	Kit
LOC_Os10g31970	SNF2 family N-terminal domain containing protein	yes	yes	yes	unknown
LOC_Os10g36703	CPuORF40 - conserved peptide uORF-containing transcript	yes	yes	yes	unknown
LOC_Os11g08200	eukaryotic aspartyl protease domain containing protein	yes	yes	yes	unknown
LOC_Os11g14160	transposon protein, putative, Pong sub-class	yes	yes	yes	unknown
LOC_Os11g28184	expressed protein	unknown	yes	yes	yes
LOC_Os11g41890*	RNA recognition motif containing protein, putative	yes	yes	yes	unknown
LOC_Os12g02250	STE_PAK_Ste20_Slob_Wnk.3 - STE kinases include homologs to sterile 7, sterile 11 and sterile 20 from yeast	yes	yes	yes	unknown
LOC_Os12g42810	mov34/MPN/PAD-1 family protein	yes	yes	yes	unknown