



Genomic dissection of the seed

Michael G. Becker^{1†}, Ssu-Wei Hsu^{2†}, John J. Harada² and Mark F. Belmonte^{1*}

¹ Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada

² Department of Plant Biology, University of California Davis, Davis, CA, USA

Edited by:

Paolo Sabelli, University of Arizona, USA

Reviewed by:

David G. Oppenheimer, University of Florida, USA

Hannetz Roschttardt, University of Wisconsin-Madison, USA

Wilco Ligterink, Wageningen University, Netherlands

*Correspondence:

Mark F. Belmonte, Department of Biological Sciences, University of Manitoba, 50 Sifton Road, Winnipeg, MB R3T 2N2, Canada
e-mail: mark.belmonte@umanitoba.ca

[†]Michael G. Becker and Ssu-Wei Hsu have contributed equally to this work.

Seeds play an integral role in the global food supply and account for more than 70% of the calories that we consume on a daily basis. To meet the demands of an increasing population, scientists are turning to seed genomics research to find new and innovative ways to increase food production. Seed genomics is evolving rapidly, and the information produced from seed genomics research has exploded over the past two decades. Advances in modern sequencing strategies that profile every molecule in every cell, tissue, and organ and the emergence of new model systems have provided the tools necessary to unravel many of the biological processes underlying seed development. Despite these advances, the analyses and mining of existing seed genomics data remain a monumental task for plant biologists. This review summarizes seed region and subregion genomic data that are currently available for existing and emerging oilseed models. We provide insight into the development of tools on how to analyze large-scale datasets.

Keywords: *Arabidopsis*, next generation sequencing, oilseed, RNA seq, seed, soybean, transcriptome

INTRODUCTION

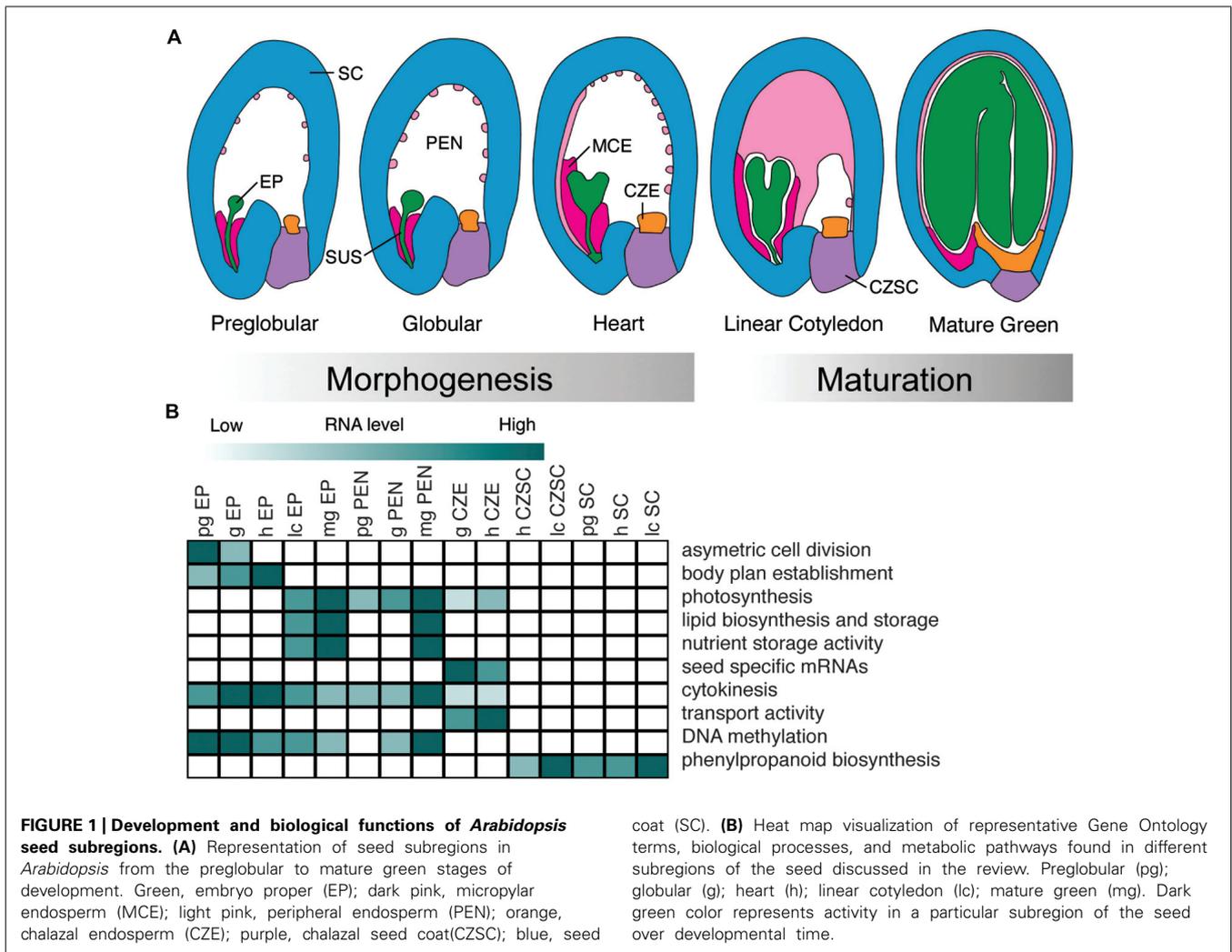
With the world population expected to reach over 9 billion by the middle of the 21st century, one of the biggest challenges facing humanity will be the production of sustainable food supplies (Godfray et al., 2010; Cleland, 2013). To accommodate world food demands, it is estimated that crop production will need to double without increasing current agricultural land use (Foley et al., 2011; Tilman et al., 2011; Ray et al., 2013). Since the direct consumption of seeds and their use as animal feed account for more than 70% of the human diet (Sreenivasulu and Wobus, 2013), recent discussions on food security have turned to enhancing crop production through seed genomics. Seed genomics is the study of genomes and the expression of genes that are required to make a seed. This includes the spatial and temporal expression and regulation of all genes active during seed development. While classical breeding strategies have proven to be effective in producing more robust and productive plant cultivars, they can be complemented and greatly improved through the utilization of genomics-based knowledge (Tester and Langridge, 2010; Feuillet et al., 2011; Langridge and Fleury, 2011).

A seed is formed upon fertilization of the female gametophyte and early stages of development involve the deterioration of maternal gametophytic structures and the establishment of the sporophyte. Seed development is initiated by a double fertilization event that results in a seed that can be divided into three distinct regions: the embryo, the endosperm, and the seed coat (SC; Ohad et al., 1999; Le et al., 2007). In the first fertilization event, a sperm, and egg cell nucleus fuse, resulting in a zygotic embryo. The embryo is part of the next sporophytic plant generation. The endosperm results from the second fertilization event between the sperm and central cell, and it will serve to support the embryo during the early stages of seed development and/or

seedling growth. Finally, the seed coat (SC) is of maternal origin and is derived from the integuments that form during ovule development. The SC transfers assimilates from the maternal plant and serves to protect the embryo throughout seed development. Further, the developmental programs that underlie seed development can be divided into two distinct phases. First, during morphogenesis, the body plan of the embryo is established and the nuclei of the endosperm proliferate. Second, during the maturation phase, large shifts in gene activity are observed across all three regions of the seed, initiating the accumulation of storage materials that help to protect the embryo in preparation for desiccation.

We can further dissect seed regions into subregions. In numerous plants including *Arabidopsis*, the zygote differentiates into the embryo proper, which will become cotyledonous and eventually form the vegetative plant, and the suspensor, which acts to facilitate communication between the embryo proper and surrounding seed regions. The endosperm develops into three distinct subregions: the micropylar endosperm (MCE, proximal to the embryo), the peripheral endosperm (PEN), and the chalazal endosperm (CZE, distal to the embryo; Brown et al., 2003). The maternally derived SC can be divided into two subregions, the chalazal seed coat (CZSC), and distal SC (Figure 1A). Depending on the model seed, these subregions can further be divided into tissue and cell types.

This review focuses on the genomic analysis of seed regions and subregions using established and emerging plant models. We discuss how genomics has been used successfully to study the development of the embryo, endosperm, and SC regions of the seed, and how new cutting-edge tools can be used to further dissect every cell and tissue of the seed into subregions for further interrogation. Finally, we present tools on how to analyze large-scale transcriptome datasets.



CHARACTERIZING THE SEED TRANSCRIPTOME

In the current genomics era we have uncovered a number of developmental and regulatory pathways responsible for making a seed. However, we still have yet to fully understand all of the mechanisms responsible for the coordination of gene activity underlying the sophisticated development of all seed regions and subregions. Many regulatory mechanisms surrounding primary and secondary metabolism, hormone regulation, gene imprinting, transcriptional-, translational-, and post-translational regulation all operate in concert to mediate the complex processes occurring during seed development. These processes are under the regulation of 100s and 1000s of genes that are often obscured by genetic redundancy and thus difficult to identify using traditional forward genetics screens (Curtin et al., 2011). Arguably, the best way to investigate coordinated events such as cell fate specification, differentiation, and morphogenesis of the developing seed is by monitoring the expression of large gene sets with high throughput genomics-focused microarray and sequencing strategies. Bioinformatic analyses can then be used to identify transcriptional networks and key regulators of seed development.

As Next Generation Sequencing experiments such as deep genomic sequencing, RNA-, small RNA-, and DNA methylome-sequencing become commonplace in the laboratory and as sequencing technologies continue to evolve, the challenge faced by the scientific community is no longer the acquisition of data, but rather compiling and analyzing the data. Publicly available databases like NCBI (The National Center for Biotechnological Information¹), GEO (Gene Expression Omnibus²), and SRA (Sequence Reads Archive³) contain large amounts of DNA microarray and nucleic acid sequence data that can be queried and mined to provide answers to challenging biological questions about the seed.

USING DNA MICROARRAYS TO PROFILE THE SEED

The Affymetrix ATH1 GeneChip microarray was one of the most widely used tools to profile the *Arabidopsis* transcriptome, and it was used to investigate numerous processes underlying

¹<http://www.ncbi.nlm.nih.gov>

²<http://www.ncbi.nlm.nih.gov/geo>

³<http://www.ncbi.nlm.nih.gov/sra>

the seed including gibberellin response (Ogawa et al., 2003), response to abscisic acid (Nishimura et al., 2007), seed dormancy (Finch-Savage et al., 2007), seed imbibition (Nakabayashi et al., 2005; Preston et al., 2009), seed germination (Dean Rider et al., 2003; Penfield et al., 2006; Dekkers et al., 2013), and development (Day et al., 2008; Le et al., 2010; Dean et al., 2011; Belmonte et al., 2013; Khan et al., 2014).

Le et al. (2010) published the *Arabidopsis* seed transcriptome at seven stages of development from ovule to seedling and identified putative regulators of seed development. At each stage of development, between 8779 and 13,722 distinct mRNAs were detected at the level of the GeneChip with 15,563 unique transcripts detected over all stages of seed development. Of these, only 2% (289) of the transcripts were considered seed-specific with the vast majority being specific to a given stage of development (e.g., globular-cotyledon). Of these seed-specific genes, 17% coded for transcription factors (TFs) and contained known regulators of seed development, such as *LEAFY COTYLEDON1 (LEC1)*, *LEAFY COTYLEDON2 (LEC2)*, *FUSCA3 (FUS3)*, and *MEDEA* (Le et al., 2010).

Similar analyses were conducted for developing soybean seed from five developmental time points ranging from mid-maturation through seed desiccation (Jones et al., 2010). This study noted an increase in TF activity late in seed development. TFs accumulating late in development included those involved in ethylene and auxin responses, as well as genes that were largely uncharacterized in soybean. Orthologous genes in *Arabidopsis* and rice suggest these genes are involved in processes such as abscisic acid and gibberellic acid signaling, sugar and nitrogen metabolism, and germination.

PROFILING THE SEED USING LASER MICRODISSECTION COUPLED WITH MICROARRAYS

Traditional studies that isolated seed regions like the embryo, endosperm, and SC for seed genomics used forceps or fine needles. The lack of precision of these manual techniques makes it nearly impossible to isolate individual regions without contamination from neighboring cells or tissues. These challenges limit the resolution of genomics research and dilute low abundant transcripts that may otherwise be detected using more sophisticated dissection methods. Regardless of the dissection tool used, the advancement of genomics-based seed research relies on contamination-free isolation of the cells and tissues of interest.

Currently, the most successful way to dissect regions and subregions of the seed for genomics studies without contamination of other cells types is through laser microdissection (LMD) technologies (Khan et al., 2014). Whole-seed mRNA profiling experiments provided some of the most informative seed genomic data across developmental time for *Arabidopsis* and soybean, but the application of LMD to these seeds provided higher resolution and more sensitive profiles of gene activity in developing seed. For example, Casson et al. (2005) dissected the *Arabidopsis* embryo to study mechanisms associated with apical / basal polarity. This study detected expression of ~65% of the 22,810 probe sets on the ATH1 array during the early

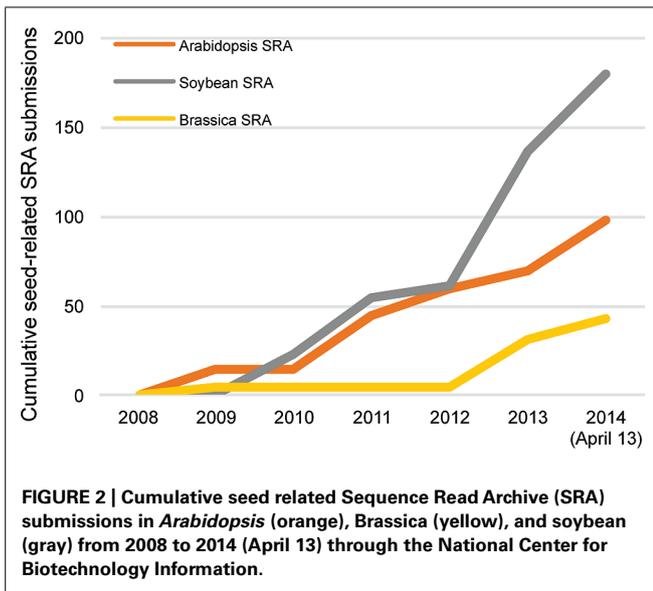
stages of embryo development. Characterization of the spatial and temporal expression of 220 genes known to cause defects in embryo development when mutated, including *PASTIC-CINO1*, *PINOID*, *PIN-FORMED3*, and *PIN-FORMED4* during embryo development provided insight into their control. Further, several of these genes are being used as markers for the embryo.

The endosperm has been a difficult seed region to study using transcriptome analysis given that the endosperm subregions are not easily isolated. LMD has proven to be an effective and contamination-free technique to isolate the individual subregions of the endosperm for transcriptional profiling (Day et al., 2008). An initial study identified 800 genes, 27 encoding TFs that are preferentially expressed during early endosperm development. Biological processes associated with the progression and control of the cell cycle, DNA processing, chromatin assembly, protein synthesis, cytoskeleton and microtubule-related processes, and cell/organelle biogenesis were all predicted to characterize endosperm proliferation and cellularization.

The most comprehensive developmental series of any seed was recently published by Belmonte et al. (2013) with the goal of identifying all of the genes and defining the gene regulatory networks responsible for guiding seed development. Thirty-six seed subregions across five developmental stages revealed complex dominant patterns of gene activity in both space and time in *Arabidopsis* (Belmonte et al., 2013; data available at seedgenenetwork.net). The combination of LMD and the ATH1 GeneChip identified at least 17,594 distinct mRNAs that are detectable during seed development and 1,316 of those mRNAs are specifically expressed in the *Arabidopsis* seed compared to vegetative and reproductive tissues. Similar data describing mRNA profiles at high spatial resolution are also available for soybean from experiments that used the Affymetrix soybean GeneChip to analyze 40 subregions across four developmental stages (Le et al., 2007; data available at seedgenenetwork.net). The reader is referred to Nelson et al. (2006) and Day et al. (2007) for reviews on methods and protocols used for LMD of plant tissues (Figure 1B).

USING NEXT GENERATION SEQUENCING TO PROFILE THE SEED

There are a number of advantages to NGS sequencing technology when compared to DNA microarrays: (i) the ability to detect low abundance transcripts, (ii) the identification of novel alternatively spliced isoforms of mRNAs, (iii) little requirement for *a priori* knowledge of the organism, (iv) increased sensitivity in the detection of differentially expressed genes, (v) more reproducible results, and (vi) the ability to compare expression profiles between distantly related organisms. For example, RNA sequencing facilitated the study of oil accumulation in four non-model oilseeds (or “emerging models”): castor (*Ricinus communis*), rapeseed (*Brassica napus*), burning bush (*Euonymus alatus*), and nasturtium (*Tropaeolum majus*; Troncoso-Ponce et al., 2011). These species differ in their location for oil deposition, triacylglycerol composition and content. Analysis of the data revealed a core set of well-conserved enzymes involved in triacylglycerol production that exhibit similar temporal expression



patterns in all species, suggesting a conserved evolutionary relationship in the production of seed oil. Putative regulators and mediators of oil production in *Arabidopsis* were identified and an online resource, “ARALIP⁴,” was established to facilitate utilization of these data. It is important to note that while NGS has several advantages over microarray technology, the detection of low abundant transcripts as well the detection of alternative splice sites is largely dependent on the depth of sequencing and should be carefully considered during the design of the experiment.

Many other RNA sequencing studies of seed genomics have focused on soybean, largely because of its global economic importance. An indication of this emphasis is that seed-related submissions of soybean RNA sequencing data to the SRA and NCBI databases nearly double those of *Arabidopsis* (Figure 2). This has produced several large datasets for soybean seed development. Two particular studies stand out, one that profiled the whole soybean seed at seven time points between 10 and 42 days after fertilization (Severin et al., 2010), and an independent study focusing on whole soybean seeds at 15–65 days after fertilization (Chen et al., 2012). These studies showed that 49,151 transcripts are detected during seed development, ~12,000 mRNAs more than the 37,500 transcripts represented on the current soybean Affymetrix array. Furthermore, 9,930–14,058 (Severin et al., 2010) and 11,592–16,255 (Chen et al., 2012) transcripts are differentially expressed compared to the earliest stage of seed development. Both of these studies provide examples of how RNA sequencing data can be mined using a range of bioinformatics approaches including gene ontology term enrichment and co-expression analyses.

Next Generation Sequencing also provides an effective method for the characterization of small RNA (sRNA) populations within the developing seed. Two classes of sRNAs highly expressed within seed tissues are microRNAs (miRNAs) and small interfering RNAs

(siRNAs). miRNAs are ~21 nucleotide, single-stranded, non-coding RNAs that mediate the degradation or translational inhibition of target mRNAs with complementary nucleotide sequences (Chen, 2012). siRNAs, derived from double-stranded RNA, cause the degradation of target mRNAs and carry out *de novo* deposition of repressive chromatin marks and will be discussed later in this review.

Much of the recent work profiling sRNAs during seed development focus on economically important emerging models. Two independent studies examined sRNA populations in soybean, focusing on the identification of miRNAs active during development and their putative targets (Song et al., 2011; Shamimuzzaman and Vodkin, 2012). Of the miRNA targets identified, 50% (Song et al., 2011) and 82% (Shamimuzzaman and Vodkin, 2012) were TFs, including auxin response factors and growth regulating factors. Eleven annotations were found in both datasets including Argonaute Protein, Auxin Response Factor, Growth Regulating Factor, HD-ZIP TF, No Apical Meristem protein, TCP Family TF, and Nuclear Factor YA. These studies also report an increase in mRNA target diversity late in development, suggesting miRNAs have a role in the shift into maturation, which agrees with data from earlier work done with *Arabidopsis* (Tang et al., 2012).

Huang et al. (2013) characterized *B. napus* sRNA populations in whole seeds at nine time points in development and in dissected endosperm, embryo, and SC at three of those stages. Similar to *Arabidopsis* and soybean, the authors suggest that miRNAs have a role in controlling seed maturation. In addition, 279 miRNAs were identified that had been previously reported, including 182 in *Arabidopsis* and 56 in soybean. Also in *B. napus*, Zhao et al. (2012) characterized miRNA populations in high- and low-oil content seeds, and they identified putative miRNA regulators of oil metabolism.

Several databases for miRNAs and their mRNA targets are available to the researcher. Currently 427, 573, and 92 mature miRNA sequences for *Arabidopsis*, soybean, and canola, respectively, have been deposited in miRBase, an online database for published miRNA sequences (⁵Kozomara and Griffiths-Jones, 2014). Another database, MiRTarBase (⁶Hsu et al., 2014) contains experimentally confirmed miRNA-target interactions (Hsu et al., 2014). In addition, MiRFANs (⁷Liu et al., 2012) stores miRNA functional annotations specifically for *Arabidopsis*, and it includes an analysis toolbox.

The production of data from Next Generation Sequencing studies is providing the scientific community with vast amounts of genomic data that can be mined to answer many important biological questions about the seed. Dramatic improvements to seed transcriptome experiments, including enhanced sequencing chemistries and better bioinformatics tools should provide the necessary tools and data required to answer these questions. With Next Generation Sequencing, subtle changes to the transcriptome can now be detected with high confidence and exploited to identify most of the genes and gene products responsible for seed development.

⁵<http://www.mirbase.org>

⁶<http://mirtarbase.mbc.nctu.edu.tw>

⁷www.cassava-genome.cn/mirfans

⁴<http://aralip.plantbiology.msu.edu/>

GENOMICS OF EMBRYO DEVELOPMENT

Embryogenesis is the developmental period during which the zygote differentiates into the mature embryo. Embryo development can be divided temporally into two phases, morphogenesis and maturation (Goldberg et al., 1994). During the morphogenesis phase, the diploid zygote derived from fertilization of the egg cell by a sperm cell undergoes an asymmetric cell division, producing the apical and basal cells (Lau et al., 2012). In many plants, the apical cell gives rise to most of the embryo proper. The basal cell develops largely into the suspensor, although the uppermost suspensor cell divides to form the hypophysis that will become the quiescent center of the root apical meristem and the central root cap cells of the embryo proper. Development of the embryo proceeds along two primary axes. Along the apical-basal axis, the embryo becomes sequentially partitioned into specific pattern elements that become the cotyledons, shoot apical meristem, hypocotyl, root, and root apical meristem. The embryo proper also becomes compartmentalized along its radial axis to generate the embryonic tissue systems: procambium, ground tissue, and protoderm. The suspensor is an ephemeral structure of the embryo that serves a structural role by pushing the embryo proper into the nutrient-rich endosperm and a physiological role by transferring nutrients and growth factors to the embryo proper at early developmental stages (Kawashima and Goldberg, 2010).

As the embryo transitions from the morphogenesis to the maturation phase, morphogenetic processes, including cell division, become largely repressed (Harada, 1997; Vicente-Carbajosa and Carbonero, 2005). During the maturation phase, the embryo acquires the ability to withstand stresses imposed by desiccation that occur late in seed development and accumulates storage proteins, lipids, and/or carbohydrates to massive amounts, causing the embryo to grow as a result of cell expansion. The storage macromolecules serve as a nutrient source for the developing seedling during post-germinative development. By the end of the maturation phase, the embryo is quiescent metabolically and arrested developmentally, and it remains in this state until conditions appropriate for germination and post-germinative development are perceived.

CONTRIBUTIONS OF THE MATERNAL AND PATERNAL GENOMES TO EARLY EMBRYO DEVELOPMENT

The zygote represents the first stage of the morphogenesis phase, and two studies have addressed the question of when the zygotic genome becomes active transcriptionally following fertilization of the egg cell. In animals, early embryonic development is regulated by maternal mRNAs deposited in the egg prior to fertilization, and the zygotic genome becomes transcriptionally active several cell cycles after fertilization (Tadros and Lipshitz, 2009). The maternal-to-zygotic transition was analyzed in *Arabidopsis* by sequencing RNAs from early stage embryos derived from crosses between plants of different ecotypes and using single nucleotide polymorphisms to distinguish mRNAs derived from maternal and paternal alleles. Autran et al. (2011) reported that the majority of mRNAs in an *Arabidopsis* embryo at the two to four cell embryo proper stage are from the maternal genome,

although approximately 10% of mRNAs are encoded by paternal alleles at this early stage. The paternal contribution to the mRNA population increased to 36% by the globular stage, which was interpreted to represent a gradual activation of the paternal genome. Paternal genome activity is maternally regulated through epigenetic mechanisms involving RNA-dependent DNA methylation, KRYPTONITE-mediated histone methylation, and CAF-1 complex-induced histone exchange (Autran et al., 2011). By contrast, a separate study of the maternal-to-zygotic transition reported that maternal and paternal genomes contribute almost equally to the transcriptomes of *Arabidopsis* embryos at the earliest stages of embryogenesis (Nodine and Bartel, 2012). Many mRNAs that are undetectable in the egg and sperm constitute the top 50% most abundant mRNAs in one or two-cell embryos, suggesting that the zygotic genome is activated immediately after fertilization and plays a major regulatory role during early embryogenesis. Discrepancies between the findings of these two studies may have resulted from the use of different *Arabidopsis* ecotypes by the two laboratories (Baroux et al., 2013). Alternatively, the high proportion of maternally derived mRNAs may have resulted from contamination of embryo samples by mRNAs from the SC that is entirely of maternal origin (Nodine and Bartel, 2012). Nevertheless, both studies demonstrated that the maternal-to-zygotic transition occurs at the earliest stage of embryo development in *Arabidopsis*.

ROLE OF microRNAs IN THE TRANSITION FROM THE MORPHOGENESIS TO MATURATION PHASE

The transition from the morphogenesis to the maturation phase represents a major shift in the developmental programs that occur during embryogenesis (Harada, 1997; Vicente-Carbajosa and Carbonero, 2005). The transcriptomes of *Arabidopsis* embryos that were isolated from the seed by LMD or hand dissection were profiled at several stages of development (Xiang et al., 2011; Belmonte et al., 2013), and these studies demonstrated that gene expression changes dramatically as embryos transition into the maturation phase. For example, the vast majority of mRNAs that accumulate in the embryo proper at a specific stage of development do so at the maturation phase. This gene set is enriched for those involved in maturation processes, including mRNAs encoding storage proteins, oilbody proteins, and proteins involved in lipid storage.

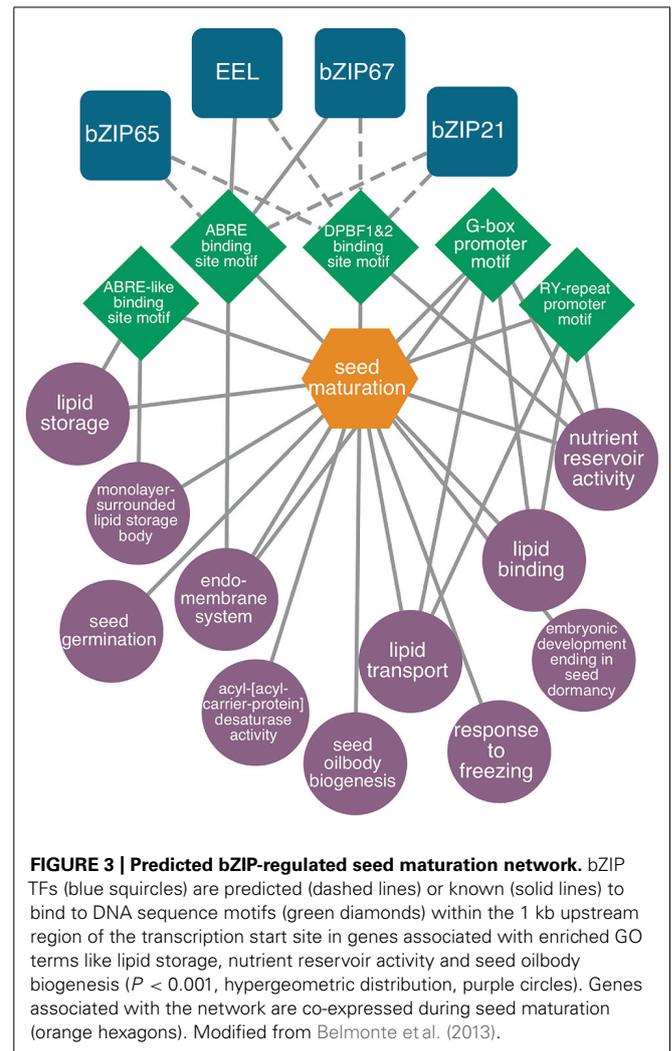
microRNAs play a critical role in controlling the transition from the morphogenesis to the maturation phase (Nodine and Bartel, 2010; Willmann et al., 2011). The role of miRNAs in controlling the transition from morphogenesis to maturation phase was revealed by studies of mutations affecting *DICER-LIKE1* (*DCL1*), which encodes an enzyme required for miRNA biosynthesis. Early in embryo development, loss-of-function *dcl1* mutants display abnormal cell division patterns in the hypophysis, a cell that will become incorporated into the root apical meristem, and in subprotodermal regions of the embryo. These findings were interpreted to suggest that miRNAs are required for embryo patterning events that occur during the morphogenesis phase (Nodine and Bartel, 2010; Willmann et al., 2011). Transcriptome analyses showed that mRNAs that normally accumulate specifically during the maturation phase, including those

encoding storage proteins, oil body proteins, lipid biosynthesis enzymes, and several transcriptional regulators of the maturation phase, accumulate prematurely in *dcl1* mutant embryos. By contrast, two TFs, ASIL1, and HDA6/SIL1, that normally repress maturation genes after germination were downregulated in *dcl1* mutants (Willmann et al., 2011). These results, along with the finding that chloroplast maturation occurs earlier in *dcl1* mutant than wild-type embryos, were interpreted to indicate that miRNAs are required to repress maturation processes during the morphogenesis phase and that the precocious onset of the maturation phase in *dcl1* mutants causes defects in pattern formation. In particular, one set of miRNAs and their target mRNAs were implicated to mediate temporal control of the maturation phase. In *dcl1* mutants, disruption of *miR156* accumulation causes the premature upregulation of two differentiation promoting TFs, SPL10, and SPL11, and experiments analyzing the effects of altering *SPL10* and *SPL11* expression suggested that they are at least partially responsible for repressing the maturation processes early in embryogenesis. A different miRNA, *miR166*, has been shown to repress genes expressed specifically during the maturation phase in vegetatively growing plants (Tang et al., 2012). Together, these observations suggest miRNAs play critical roles in controlling embryonic processes.

MATURATION GENE REGULATORY NETWORKS

Several studies have focused on understanding the gene regulatory networks that operate during the maturation phase of seed development (reviewed by Santos Mendoza et al., 2005; Gutierrez et al., 2007; Braybrook and Harada, 2008; Holdsworth et al., 2008; Suzuki and McCarty, 2008; Junker et al., 2010). To gain insight into embryo maturation gene regulatory networks, Belmonte et al. (2013) identified DNA sequence motifs that are overrepresented in the 5' flanking regions of a set of genes that are expressed in embryos specifically during the maturation phase. TFs that are known or predicted to bind these overrepresented DNA sequence motifs were also identified, permitting a putative gene regulatory network to be created. The network included a number of *cis*-acting DNA elements that have been shown previously to regulate genes expressed during the maturation phase, including the *ABRE*, *ABRE-like*, *DPBF1*, *DPBF2*, and *RY* motifs. Identified among the TFs known to bind these motifs were EEL and bZIP67, which are known to regulate genes during the maturation phase. An example of a maturation gene regulatory network is presented in **Figure 3** and a description of the construction of gene regulatory networks is presented in "Identifying regulatory networks required to program the *Arabidopsis* seed" below.

Studies to characterize regulators of the maturation phase have focused on the *Arabidopsis* LEC1, LEC2, FUS3, and ABI3 TFs (Koornneef et al., 1984; Meinke, 1992; Keith et al., 1994; Meinke et al., 1994; West et al., 1994). LEC1 is a HAP3 (a.k.a. NF-YB) subunit of the CCAAT-binding (NF-Y) TF (Lotan et al., 1998), whereas LEC2, FUS3, and ABI3 are B3-domain TFs (Giraudat et al., 1992; Luerssen et al., 1998; Stone et al., 2001). The central roles of these maturation TFs in controlling embryo and seed development was established initially through investigations of mutations in these genes. Loss-of-function mutations



in these maturation TF genes cause embryo lethality or the ablation of embryo parts, because mutant embryos are intolerant of desiccation and storage protein and lipid accumulation is defective. Ectopic expression of these maturation TF genes induces somatic embryo development, fatty acid biosynthesis, oil body accumulation and storage protein biosynthesis in vegetative cells (Parcy et al., 1994; Lotan et al., 1998; Kagaya et al., 2005a; Santos Mendoza et al., 2005; Mu et al., 2008; Stone et al., 2008; Feeney et al., 2013).

The maturation TFs LEC1, LEC2, FUS3, and ABI3 are involved in complex and redundant regulatory interactions during embryo development (reviewed by Braybrook and Harada, 2008; Junker et al., 2010). Genetic and molecular experiments have shown that LEC1 functions upstream of LEC2, FUS3, and ABI3 and, therefore, is likely to act at or near the top of the regulatory hierarchy controlling maturation (Kagaya et al., 2005b; To et al., 2006). Redundancy is observed in interactions among the other maturation TFs that is dependent on their spatial location in the embryo (To et al., 2006). For example, the *FUS3* gene is regulated by LEC1, LEC2, and ABI3 in cotyledons, by LEC2 and ABI3 in the embryonic axis, and by LEC2 and FUS3 in the root tip. Together, the results

suggest that these maturation TFs play key but complex roles in the regulatory network controlling the maturation phase of seed development.

In recent years, initial dissection of the maturation gene regulatory network has occurred through the genome-wide identification of target genes that are directly regulated by the maturation TFs. Direct target genes are generally defined as those that are bound by a TF, as determined by chromatin immunoprecipitation experiments, and that are regulated by that TF. Genes that are up- and downregulated by a TF are often identified by comparing their mRNA levels in embryos with a mutation in the TF gene versus wild type. Alternatively, regulated genes are identified by using inducible forms of the TF. Imposing a gene expression constraint on the identification of direct target genes is important, because fewer than 10% of genes that are bound by a TF are regulated by that TF (Farnham, 2009). Genome-wide analysis identified 98 genes that are both bound by ABI3 and regulated following the induction of ABI3 activity, including genes encoding 2S seed albumins, 12S seed storage globulins, oleosins, and desiccation-related LEA proteins (Monke et al., 2012). Most of these target genes are generally expressed during the maturation phase, and they require abscisic acid for their activation, consistent with the observation that mutations in *ABI3* confers insensitivity to ABA (Koornneef et al., 1984). Analysis of the *ABI3* target genes identified two DNA sequence motifs that are both overrepresented in the first 250 bp upstream of the transcription start site: a RY element that is known to be bound by ABI3 and a G-box motif. The G-box is part of a well-characterized ABA-responsive element (e.g., *ABRE*) motif that interacts with bZIP TFs. These findings are consistent with previous studies showing that ABI3 interacts with a bZIP TF to regulate the transcription of genes involved in maturation processes (Nakamura et al., 2001; Lara et al., 2003).

Target genes for another B3-domain maturation TF, FUS3, were identified from embryonic culture tissue overexpressing the *AGL15* gene that expresses *FUS3* constitutively (Wang and Perry, 2013). FUS3 target genes were enriched for maturation processes, and showed a 17% overlap with *ABI3* target genes. The 5' flanking regions of the FUS3 target genes were enriched for RY and G-box motifs. These studies confirmed on a genome-wide scale that there is at least partial redundancy in the functions of FUS3 and *ABI3*. FUS3 also directly regulates another B3-domain TF, VAL1, which along with VAL2 and VAL3, acts as repressors of the maturation network during seedling development (Suzuki and McCarty, 2008). FUS3 was also shown to regulate miRNA genes, including *miR156*, *miR160*, *miR166*, *miR169*, *miR369*, and *miR390*. Thus, FUS3 may be involved in controlling the shift from the morphogenesis to maturation phase given the proposed role of *miRNA156* in this transition.

Genetic and molecular studies place *LEC1* at or near the top of the regulatory hierarchy controlling the maturation phase (Kagaya et al., 2005a; To et al., 2006). Analysis of genes that are bound and regulated by *LEC1* identified two genes, *LEC1-LIKE* and *FATTY ACID BIOSYNTHESIS2*, which suggested a potential role for *LEC1* in lipid biosynthesis and other maturation processes (Junker et al., 2010). Other direct target genes regulated by *LEC1* are involved in auxin and brassinosteroid biosynthesis and signaling, light

responses and transcription regulation. The studies also demonstrated an interaction between *LEC1* and ABA signaling. For example, although *LEC1* can bind to the 5' flanking sequences of the *YUC10* gene that encodes an auxin biosynthetic enzyme in the absence of ABA, *LEC1*-induced *YUC10* expression is ABA dependent. Together, these results suggest that *LEC1* plays an integrative role during plant development.

GENOMICS OF ENDOSPERM DEVELOPMENT

Endosperm development is initiated with the fertilization of the central cell of the female gametophyte by a sperm cell and proceeds through three distinct stages in most angiosperms: syncytial, cellularization, and cellular (Olsen, 2004; Li and Berger, 2012). During the syncytial stage, the endosperm undergoes nuclear divisions without corresponding cell divisions, generating a syncytium of nuclei that each associates with a cytoplasmic region to form nuclear-cytoplasmic domains (Brown et al., 1999). This period of syncytial development is followed by cellularization in which cell walls form around nuclear cytoplasmic domains, beginning after the eighth nuclear divisions in *Arabidopsis*. Cellularization proceeds in a wave-like manner from the micropylar to the chalazal ends of the endosperm (Figure 1A). During the cellular stage, additional endosperm cells are formed through cytokinesis primarily at the periphery of the endosperm. Complex patterning of the endosperm is perhaps best exemplified by the *Brassicaceae*, including *Arabidopsis* and canola, in which three distinct endosperm subregions form corresponding to their positions within the seed: micropylar, peripheral, and chalazal (Figure 1A). These spatial domains are specified at the earliest stage of endosperm development in that their nuclear, cytoskeletal, and cytoplasmic characteristics and positions within the endosperm are distinguished by the fourth mitotic division (Brown et al., 2003). Depending upon the species, the endosperm remains largely intact throughout seed development as occurs in cereal grains, or it degrades as in *Arabidopsis*, canola, and soybean seeds.

ENDOSPERM DOMAINS HAVE DISTINCT AND OVERLAPPING FUNCTIONS

Transcriptome analyses of the *Arabidopsis* endosperm have provided novel insights into the relationship between the micropylar, peripheral, and CZE subregions. Previous work using LMD to profile endosperm mRNA populations provided the first characterization of gene expression genome-wide in the micropylar, peripheral, and chalazal subregions (Belmonte et al., 2013). These studies showed that a small subset is expressed specifically in each endosperm subregion at virtually all stages of development, suggesting strongly that each subregion fulfills a unique function within the seed. In particular, the CZE has the largest number of genes that are expressed specifically in a single subregion of the seed and the most seed-specific genes among all subregions. Analyses of these CZE-specific genes showed that they encoded rate-limiting enzymes involved in the biosynthesis of the hormones gibberellic acid, abscisic acid, and cytokinin (Day et al., 2008; Belmonte et al., 2013), confirming the work of others who localized these enzymes to the CZE (Miyawaki et al., 2004; Lefebvre et al., 2006; Hu et al., 2008). Chalazal endosperm-derived abscisic

acid, cytokinin, and gibberellic acid, respectively, are involved in controlling seed dormancy, endosperm cellularization, and growth of maternal tissues. Thus, the CZE may serve as a hub that supplies hormones to regulate developmental processes in developing seeds.

Analyses of the transcriptome datasets uncovered dominant patterns of gene activity for mRNAs that are involved in processes critical for seed development and that occur in all three endosperm domains and in the embryo. Clustering analyses identified a number of different gene sets that are expressed at early stages of seed development in the embryo and micropylar and PEN, but their expression in the CZE is delayed until the late developmental stages (Belmonte et al., 2013). One set encodes proteins involved in cytokinesis, consistent with the observation that embryo cells undergo cytokinesis concurrently with mitosis, whereas endosperm cellularization proceeds from the micropylar to the chalazal ends of the endosperm. Another set is involved in photosynthesis and carbon metabolism, a surprising result given that these processes were known to occur in the embryo but much less was known about their role in the endosperm. Additional analyses provided strong evidence that maturation processes occur not only in the embryo but also in all endosperm subregions. Together, these results emphasize a strong degree of overlap in gene expression programs between the embryo and endosperm regions of the seed.

GENOMIC IMPRINTING AND THE CONTROL OF SEED SIZE

The endosperm has a profound influence on seed size. It has been shown or hypothesized that the size of the endosperm early in seed development, the timing of cellularization of endosperm cells, the provisioning of maternally derived nutrients from the endosperm to the embryo, and the influence of the endosperm on the proliferation and elongation of SC cells are major determinants in specifying seed size (Scott et al., 1998; Garcia et al., 2003, 2005; Melkus et al., 2009; Ohto et al., 2009). The endosperm influences seed size through parent-of-origin effects. Parent-of-origin effects are exemplified by genetic crosses between plants of different ploidy levels. Progeny from interploidy crosses that have an excess of maternal genomes (e.g., tetraploid female crossed with diploid male) produce seeds that are smaller than self-fertilized diploid plants, whereas plants with an excess of paternal genomes (e.g., diploid female by tetraploid male) produce larger seeds (Scott et al., 1998). The parental conflict theory has been proposed to explain the antagonistic influences of the mother and father. It is hypothesized that in polygamous organisms, the father will attempt to enhance the allocation of maternally derived resources specifically to his offspring to maximize their growth, whereas the mother will try to distribute resources equally to all offspring to equalize their growth (Haig and Westoby, 1989).

Parental influences on seed size are thought to be mediated by genomic imprinting. Imprinted genes are expressed following fertilization predominately from either the maternal or paternal alleles unlike the vast majority of genes that are expressed nearly equally from both alleles. Imprinted genes are thought to control resource allocation to the embryo and therefore support its growth. Consistent with this hypothesis, an imprinted gene has

been shown to be involved in controlling maternal nutrient uptake and seed biomass (Costa et al., 2012). Imprinted genes have been identified using RNA sequencing experiments in which *Arabidopsis* plants of different ecotypes were crossed, and mRNAs from maternal and paternal alleles in the progeny were distinguished based on single nucleotide polymorphisms (Gehring et al., 2011; Hsieh et al., 2011; Wolff et al., 2011). These studies identified between 60 and 208 imprinted genes and showed that maternally expressed imprinted genes (MEGs) are more prevalent than paternally expressed imprinted genes (PEGs). Although these studies rarely identified any genes as being imprinted in the embryo, a recent study by Raissig et al. (2013) identified 11 MEGs and one PEG in the *Arabidopsis* embryo.

Genomic imprinting is regulated through epigenetic mechanisms involving DNA methylation and the Polycomb Repressive Complex 2 (PRC2). 5'-Methylcytosine in DNA is an epigenetic mark that is often associated with transcriptionally silenced genes, and PRC2 mediates gene silencing through the trimethylation of lysine 27 of histone H3 (H3K27me3, Kohler et al., 2012). To dissect the mechanisms regulating imprinted genes, Hsieh et al. (2011) and Wolff et al. (2011) analyzed the effects of mutations that cause defects in DNA methylation, DNA demethylation, and the PRC2 complex on gene imprinting. Collectively, their results showed that the DNA methylation status of MEGs correlated strongly with their imprinting. During female gametophyte development, the genome of the central cell, that is the maternal precursor of the endosperm, becomes hypomethylated globally due to the activity of DME, a DNA glycosylase that removes methylcytosine residues from DNA (Gehring et al., 2009; Hsieh et al., 2009). Hypomethylation of MEGs in the central cell results in the expression of maternal alleles of MEGs in the endosperm, whereas the paternal alleles retain their DNA methylation marks and remain silenced. The paternal alleles of some MEGs have also been shown to be silenced through the PRC2 pathway. By contrast, the paternal alleles of PEGs are active, but the maternal alleles are silenced predominately through the PRC2 pathway. These studies support the idea that demethylation of the maternal allele of some PEGs is required to permit the gene to be silenced by the PRC2 (Weinhofer et al., 2010). Thus, a complex set of epigenetic regulatory mechanisms underlies genomic imprinting.

A potential causal link between parent-of-origin effects and endosperm size came from studies of 24 nucleotide p4 siRNAs in developing endosperm (Lu et al., 2012). p4 siRNAs, which in endosperm are derived specifically from the maternal genome, function in RNA-dependent DNA methylation to target specific loci for methylation (Mosher et al., 2009; Law and Jacobsen, 2010). p4 siRNAs primarily target transposable elements for DNA methylation. However, a significant fraction of genes are closely associated with transposons, and methylation of some of these transposons influences the expression of the linked gene. Genome-wide profiling of siRNAs in interploidy crosses of *Arabidopsis* showed that 24 nt siRNAs corresponding to specific genomic loci were strongly overrepresented in endosperm of seeds with a maternal genome excess relative to seeds with a paternal genome excess. Several of these loci corresponded to genes encoding AGL TFs, one of which has been shown to inhibit endosperm cellularization

(Kang et al., 2008). These findings were interpreted to indicate that p4 siRNAs targeting AGL TFs are overrepresented in endosperm with a maternal genome excess, causing premature repression of the expression of AGL genes and precocious cellularization, resulting in a smaller seed. Together, these findings indicate a critical role for the endosperm in several aspects of seed development.

GENOMICS OF SEED COAT DEVELOPMENT

Compared to the embryo and endosperm, the SC has received little attention at the genomics level. The maternally derived SC is responsible, in part, for the evolutionary success of the seed, and it plays an integral role in filling (Verdier et al., 2013), protection, and dispersal of seeds (Haughn and Chaudhury, 2005). The SC region, like the embryo and endosperm, can further be divided into subregions based on morphological and anatomical features. For example, in *Arabidopsis* and canola, the distal SC comprised the inner and outer integuments, undergoes dramatic anatomical transformations including cell expansion, changes in cell wall deposition, and anthocyanin and mucilage accumulation followed by programmed cell death, all in preparation for seed dormancy. Conversely, the CZSC, located proximal to the funiculus, is found at the junction with the maternal plant. In seeds of legumes, like soybean, a total of six subregions have been identified: (i) endothelium, (ii) hour glass, (iii) palisades, (iv) parenchyma, (v) epidermis, and (vi) hilum. The hilum in soybean is considered to be similar in function to the CZSC in *Arabidopsis* and presents the first point of entry of material destined for filial seed compartments.

While the development and anatomy of the SC in oilseeds, such as *Arabidopsis*, soybean, and canola, have been extensively studied (Beeckman et al., 2000; Western et al., 2000; Windsor et al., 2000; Macquet et al., 2007; Young et al., 2008; Dean et al., 2011) there is remarkably little information about the genes and gene regulatory networks underlying this multicellular structure. Even less information is available about the genomics of SC development in emerging model crop systems. Of the few studies that have examined the SC at the genomics level (Jiang and Deyholos, 2010; Dean et al., 2011; Belmonte et al., 2013; Khan et al., 2014), data suggest the SC is more similar to maternal tissues than to the embryo or the endosperm. Despite the vast amount of data currently being generated and the different technologies being employed to study the SC, it is still unclear how many genes are active in each subregion and how those numbers change between species.

TRANSCRIPTIONAL REGULATION IN THE SEED COAT

When comparing the SC to other seed regions, *Arabidopsis* is the best plant model studied to date. Hierarchical clustering of GeneChip data showed differences between each subregion of the SC. It is clear that global similarities and differences exist in the SC region compared to the embryo and endosperm and have likely evolved over time to protect the embryo and to adapt to environmental conditions (Debeaujon et al., 2000). Quantitative differences in gene activity within subregions of the SC provided insight into the biological processes underlying its development. Dominant patterns of gene expression were identified from comprehensive RNA profiling of *Arabidopsis* seed subregions. This

analysis identified sets of genes that show spatial (between different subregions) and temporal (across seed development) differences in expression (Belmonte et al., 2013; Khan et al., 2014). Co-expressed gene sets were shown to represent biological processes associated with the development of SC color (Zhang et al., 2013), anthocyanin deposition (Debeaujon et al., 2003), and mucilage accumulation (Western et al., 2001), which have been extensively studied using forward genetic analyses. These studies revealed essential processes associated with the SC that are controlled by individual genes or small sets of genes, yet it was still unclear how all of these processes may be coordinated over the lifecycle of the seed.

Cellular processes that occur in the SC have been independently shown to be controlled by TFs belonging to MYB (Nesi et al., 2001; Penfield et al., 2001), HD-Zip (Johnson et al., 2002; Ishida et al., 2007), and MADS-Box (Nesi et al., 2002; Huang et al., 2011) families. Our comprehensive SC transcriptome analysis identified all of these TF mRNAs in a single analysis (Khan et al., 2014). Not only were all of these known regulators identified in our experiment, we also identified a number of possible gene targets responsible for cell fate specification, the accumulation of mucilage, the deposition of anthocyanin, flavonoid biosynthesis, and SC color.

TRANSCRIPTIONAL REGULATION OF SEED COAT COLOR

Seed coat color is an agroeconomically important trait and is determined by the presence or absence of flavonoids, more specifically, proanthocyanidins. Flavonoids are secondary metabolites produced in plants derived from the phenylpropanoid pathway and are thought to have a number of functional roles, including photoprotection (Agati et al., 2013) and cellular signaling (Pourcel et al., 2013). Proanthocyanidins accumulate exclusively in the SC. When cells in the SC die, the proanthocyanidins oxidize and polymerize to form brown pigments that darken the seed. Mutants that have defects in proanthocyanidin production form a lighter colored or transparent SC (yellow/green). The yellow/green SC coloration is often associated with other desired agroeconomic traits such as thinner SCs, decreased fiber, and higher protein and oil contents (Simbaya et al., 1995; Lipsa et al., 2011; Jiang et al., 2013b). Proanthocyanidin deficient mutants do not appear to have any major physiological disturbances other than SC color; however, some evidence suggests the mutants may have diminished responses to abiotic/biotic stress (Pourcel et al., 2013), longevity, and germination (Dean et al., 2011; Jiang et al., 2013a).

Seminal work in the genetics and biochemistry of SC color in *Arabidopsis* revealed complex networks of genes and gene products responsible for this trait (Yu, 2013). In canola and soybean, genes that contribute to SC color are more difficult to identify genetically due to redundancies within the genome. RNA sequencing of brown- and yellow-coated *B. juncea* revealed three dihydroflavonol reductase genes and three anthocyanin reductase genes that were highly expressed in the brown-seeded variety with almost no detectable expression in the yellow-seeded variety (Liu et al., 2013a). The expression of three phenylpropanoid biosynthetic genes, ten flavonoid biosynthetic genes and four regulatory genes were studied using qRT-PCR at seven

developmental stages in yellow- and brown-seeded *B. napus*. Two propanoid biosynthetic genes (*PHENYLALANINE AMMONIA LYASE*, *TRANS-CINNAMATE 4-MONOOXYGENASE*), two flavonoid biosynthetic genes (*TRANSPARENT TESTA4*, 6), five anthocyanidin/proanthocyanidin biosynthetic genes (*3,4-DICHLOROPHENOL GLYCOSYLTRANSFERASE 2*, *TRANSPARENT TESTA3*, 10, 12, 18), and three TFs (*TRANSPARENT TESTA8*, *TRANSPARENT TESTA GLABRA1*, 2) had different expression patterns in yellow seeds (Qu et al., 2013). Further, eleven quantitative trait loci mediating SC color and fiber content were identified using high-density SNP arrays in canola (Liu et al., 2013a). Together genomics studies of SC color provide new targets for improving desirable traits, such as seed oil quality, and highlight the genetic complexity of SC color (Liu et al., 2013b). The analysis and identification of new QTLs combined with RNA sequence data should provide the information needed to design improved breeding strategies.

TRANSCRIPTIONAL REGULATION IN THE CHALAZAL SEED COAT

While the distal SC has been the primary focus of numerous functional studies, the CZSC has not been studied in the same detail. Bioinformatic analysis of CZSC mRNA populations uncovered a number of transport processes that showed dynamic programs of activity across development. These processes had not been described previously because of the inaccessibility of the CZSC within the seed for experimental analysis. For example, genes associated with phloem unloading including *SUCROSE-PROTON SYMPORTER 2* (*SUC2*) and a complement of *SWEET* genes encoding sucrose efflux transporters, amino acid transport genes including *BIDIRECTIONAL AMINO ACID TRANSPORTER 1* (*BAT1*), *AMINO ACID PERMEASE 2* (*AAP2*), water transport genes encoding tonoplast intrinsic proteins (*TIP1;1*), and plasma membrane intrinsic protein are all expressed in the CZSC. These findings support the hypothesis that transport processes are enriched in the CZSC. Co-expression networks generated from transcriptome data provided insight into the regulation of these transport processes. A putative G-box regulated network controlling water and sugar transport in the developing seed through bZIP TFs, including bZIP25, bZIP28, and LRL1 (Khan et al., 2014). Functional characterization of these transcriptional regulators predicted to be associated with CZSC function presents a new avenue of targeted seed improvement through modification of maternally derived subregions.

IDENTIFYING REGULATORY NETWORKS REQUIRED TO PROGRAM THE *Arabidopsis* SEED

To better understand the underlying transcriptional mechanisms required to program the seed, an integrative systems biology approach should be applied that incorporates molecular and computational biology. First, large-scale datasets are required for such an approach, and excellent sources of seed genomic data are available at databases such as GEO and NCBI as discussed previously. However, mining this data effectively requires the development of more advanced and user-friendly tools that are available to a broader scientific audience through online

databases. Tools from the BioArrayResource⁸, Genevestigator⁹, and The *Arabidopsis* Information Resource¹⁰ are all excellent resources for genomics-based data including but not limited to whole seed, seed region, and seed subregion datasets. In addition, the seedgenenetwork.net database houses whole seed, seed region, and seed subregion transcriptome, sRNA, and DNA methylome datasets from *Arabidopsis* and soybean. Although usability of online tools continues to improve, it remains difficult to identify genes with key roles in seed development with these online tools.

Using high-resolution seed datasets from *Arabidopsis* (Le et al., 2010; Belmonte et al., 2013; Khan et al., 2014), we developed a user-friendly bioinformatics program to identify transcriptional circuits from large-scale datasets at every stage of the seed life-cycle¹¹. We identified genes, focusing our attention on TFs that are predicted to control biological processes across developmental time or that are specific to a seed subregion, including the embryo proper, micropylar endosperm, CZE, or the distal and CZSCs. The transcriptional module analysis is based on the association of a specific set of co-expressed genes with their enriched Gene Ontology terms, known DNA sequence motifs, metabolic processes, and TF families and presents the user with possible gene targets regulating biological processes within the seed.

For example, we identified a transcriptional module consisting of genes expressed specifically in the micropylar endosperm and that are enriched for the WRKY DNA sequence motif in their 5' flanking regions. Our model predicts *MINISEED3* to control processes associated with the endomembrane system in the early stages of seed development. While *MINISEED3* has previously been shown to localize to the micropylar endosperm (Luo et al., 2005), the model allows us to predict gene targets of this TF which were previously unknown (Figure 4A). We also studied a putative transcriptional network underlying the CZE. Up until recently, genetic information about this understudied subregion was lacking. However, through our integrative bioinformatics approach we identified a putative *CIRCADIAN CLOCK ASSOCIATED1*-regulated transcriptional circuit controlling ubiquitin-dependent protein catabolic processes (Figure 4B). Within the SC, we identified a number of regulators that have been previously associated with SC development, allowing a high degree of confidence in our predictive transcriptional modules (Figure 4C). The *TRANSPARENT TESTA GLABRA* complex is implicated in the regulation of flavonoid biosynthesis, and several MYB TFs (including MYB5) are implicated in the regulation of mucilage biosynthesis and the differentiation of the outer integuments (Khan et al., 2014).

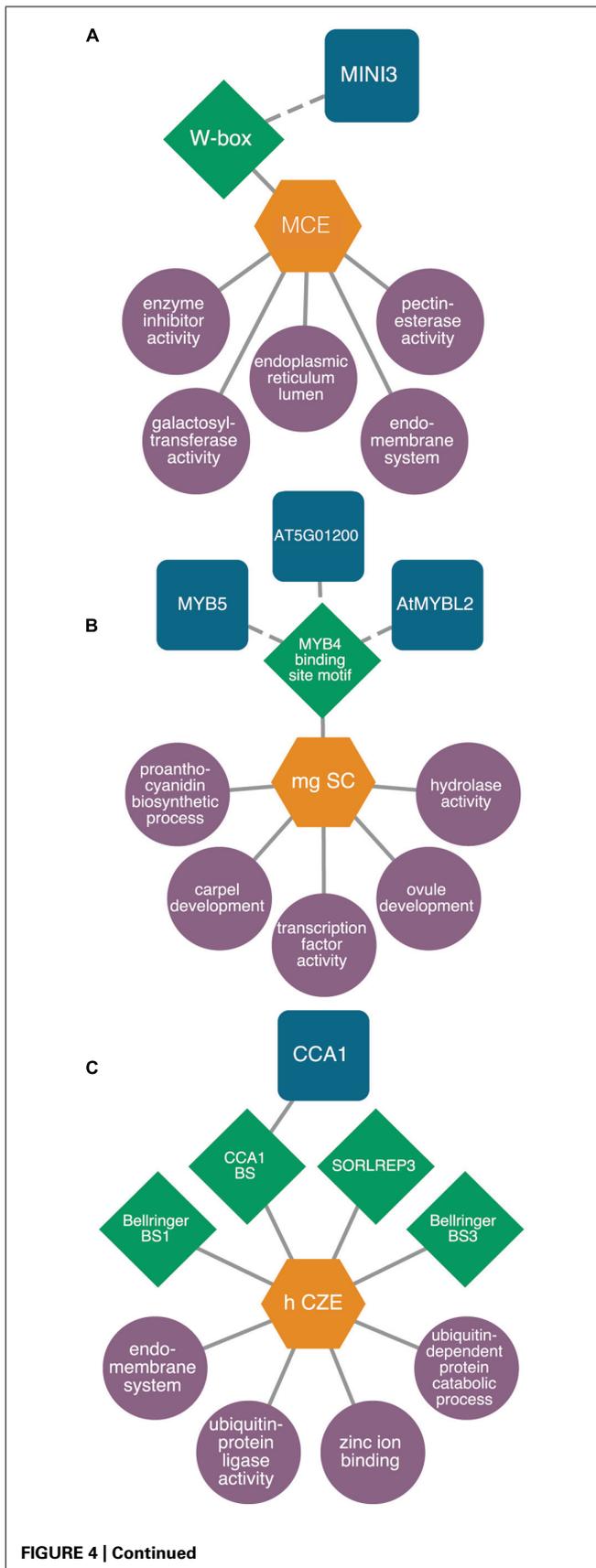
While this type of data analyses has been used successfully to identify existing transcriptional circuits, the real power of this approach lies in the identification of unknown interactions and prediction of the biological processes controlled by a TF. One of the caveats to this method is that a well-annotated genome must

⁸www.bar.utoronto.ca

⁹www.genevestigator.com

¹⁰www.arabidopsis.org

¹¹<http://seedgenenetwork.net/presentation#software>

**FIGURE 4 | Continued****Predictive transcriptional circuits in subregions of the *Arabidopsis* seed.**

(A) *MINISEED3* (*MINI3*)-W-box transcriptional circuit in the micropylar endosperm (MCE) regulating processes like the endomembrane system. (B) A *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*) module in the chalazal endosperm (CZE) of heart-stage seeds. (C) A *MYB* transcriptional module in the mature green (mg) seed coat (SC) predicted to control processes like proanthocyanidin metabolism and ovule and carpel development. TFs (blue squircles) are predicted (dashed lines) or known (solid lines) to bind to DNA sequence motifs (green diamonds) within the 1 kb upstream region of the transcription start site in genes associated with enriched ($P < 0.001$, hypergeometric distribution) GO terms (purple circles) within patterns of co-expressed gene sets (orange hexagons). All networks are modified from Belmonte et al. (2013).

be available as a reference. Thus, one of the challenges in emerging crop systems will be the annotation of genomes for which genomics research is still in its early stages. While we are beginning to understand some of the molecular mechanisms underlying the development and properties of different seed subregions and regions, the interconnectedness of these transcriptional circuits will remain a priority in the effort to elucidate the complex regulatory pathways responsible for seed development. The spectacular increase in genomic resources applicable to the seed will enable a more comparative approach to uncover and study both conserved and unique transcriptional circuits among related seed species such as the *Brassicaceae* or the *Leguminosae*. Current efforts are directed at implementing and developing computational programs to identify gene regulatory networks for important crop species like canola and soybean. The ability to predict transcriptional circuits in cell and tissue types previously thought to be inaccessible within the seed provides unprecedented insight into the regulation of biological processes over developmental time.

IDENTIFICATION OF TFs ESSENTIAL FOR SEED DEVELOPMENT

Analysis of putative gene regulatory networks is an excellent way to identify possible regulators of seed development. However, experimental validation and functional characterization of the TFs are required to validate the network. Identification of essential seed genes is a cumbersome task yet remains a priority for those interested in studying seed biology and genomics. While research has focused on essential seed genes that when mutated cause a seed lethal phenotype, other mutant phenotypes may result in defects in metabolic pathways or biochemical processes, cellular development, morphology, or other more subtle molecular phenotypes. Through our work, we have identified a number of region- and subregion-specific TFs; however, the vast majority of mutant alleles of these regulators failed to show a seed lethal phenotype (Le et al., 2010; Belmonte et al., 2013). Thus, the function of most subregion-specific TF mRNAs discovered in our work remains unknown.

Much has been learned about the seed through the use of forward genetics. Forward genetics involves generation of random mutations within an organism through radiation-, chemical-, or insertion-induced mutagenesis followed by screening for an aberrant phenotype. Systems for phenotyping mutants are

becoming increasingly automated (Fiorani and Schurr, 2013), and NGS strategies are being used to map the mutation site in what is being referred to as “fast-forward” genetics (Schneeberger and Weigel, 2011). Through forward genetics, an extensive collection of *Arabidopsis* T-DNA mutants is available through the SALK Institute (Alonso et al., 2003), and a database of essential seed genes has been established at seedgenes.org (Meinke et al., 2008) and the *Arabidopsis* Biological Resource Center.

As we continue to characterize the seed genome, forward genetics becomes increasingly ineffective as the likelihood of discovering previously uncharacterized mutants decreases. Molecular tools such as RNA interference and over-expression lines have provided researchers with important information about their genes of interest. However, new genome editing techniques utilizing the CLUSTERED REGULARLY INTERSPACED SHORT PALINDROMIC REPEATS (CRISPR)/CRISPR-Associated System (CAS; Xie and Yang, 2013), Transcription Activator-Like Effector Nucleases (TALENs; Christian et al., 2013), and Zinc Finger Nucleases (ZFNs; Zhang et al., 2010; de Pater et al., 2013), are becoming popular alternatives to classical mutagenesis. Unlike the previous approaches that relied solely on chance, emerging technologies provide an efficient means to achieve targeted mutagenesis and target multiple alleles simultaneously (Curtin et al., 2011). In addition, there is the potential for targeting non-coding regions of the genome to elucidate regulatory functions of nucleic acid sequences (Gaj et al., 2013). Of these systems, the most recent to emerge is the CRISPR/CAS system. Unlike ZFNs and TALENs that rely on complicated protein–DNA interactions, the CRISPR/CAS system uses guiding RNAs and simple base pairing between the RNA construct and target site. In addition, this technology has the ability to perform multiple genome edits by targeting more than one location simultaneously (Cong et al., 2013). This technology is also proving to have several additional practical applications, such as the modification of gene expression *in vivo* through gene fusion to transcriptional activation or repression domains (Bikard et al., 2013) or for the labeling of individual chromosomal loci (Chen et al., 2013). Taken together, the ability to manipulate transcriptional networks and fine-tune gene expression would prove valuable tools for the molecular dissection and engineering of seeds.

FUTURE DIRECTIONS

It is an exciting time to study the underlying mechanisms of seed development through genomics. The complex morphological and metabolic transformations of the seed lend themselves to intensive genomic interrogation. While seminal work dissecting cells, tissues, and organs of both *Arabidopsis* and soybean seeds has revealed an incredible abundance of information, there are still pressing questions when it comes to the coordination and regulation of seed development at the cellular and tissue levels. To answer these questions seed biologists are using modern sequencing strategies. The incredible amount of information produced by these technologies is overwhelming, and the information extracted from these analyses will only continue to improve as we perfect the chemistries and foster new collaborations with mathematics, statistics and computer science. These

in-depth analyses yield significant information about the transcriptional circuitry underlying complex tissue systems responsible for the development of the seed. Moreover, identification of transcriptional regulators from large-scale datasets will provide the necessary starting point for research focusing on improving seeds.

To achieve these goals, plant biologists are coupling cutting-edge technologies that are capable of dissecting or isolating individual cells and tissues of the seed with sequencing platforms. In addition to mRNA profiling, LMD has been coupled to genomics strategies such as bisulfite sequencing to study global changes in DNA methylation marks, degradome sequencing to study miRNA cleavage sites, and ChIP sequencing to identify protein/TF DNA interactions during seed development. DNA sequencing, bisulfite sequencing, RNA and small-RNA sequencing, degradome sequencing, ChIP sequencing, and CLIP sequencing (protein–RNA interactions) each provide a piece to the developmental puzzle, and sophisticated integrative computational analyses will be required to put all of the pieces together. Thus, the development of integrative computational tools to analyze complex and possibly disparate datasets in all plants will remain a major challenge for the scientific community.

Despite the tremendous advances in genomics-focused research including NGS platforms and the continuing reduction in the cost and production of high-resolution datasets, functional characterization of genes responsible for seed development, especially in emerging model systems, remains a challenge. Functional testing and characterization of the biological information derived from the billions of data points that sample the dynamic biological processes underlying seed development will take decades using current molecular biology tools. Thus, high-throughput functional characterization of genes and gene products remains a top priority for plant biologists. There are four areas of seed genomics and its application that we suggest need to be targeted to further improve our understanding of the seed: (i) update and curate small- and large-scale genomics data in publicly available databases; (ii) implement user-friendly data analysis pipelines and educate scientists on how to use them effectively; (iii) profile and characterize the genomes of emerging models important for global crop production and development; (iv) functionally characterize every gene responsible for plant traits relevant to sustainable agriculture.

CONCLUSION

Current advancements in seed genomics are illuminating the genetic forces driving seed development. It is now possible to identify most of the genes responsible for guiding seed development in every cell, tissue, and organ throughout the seed lifecycle. Together, modern breeding strategies that include information derived from genomics-based research will provide the necessary tools to improve seeds: seeds with improved nutritional value, that can endure adverse environmental conditions, or one that can withstand biological attack. Our dependence on seeds for food, fuel, and other resources means seed improvement research through genomics will continue to have a significant impact on global biosustainability.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the Plant Genome Program of the National Science Foundation to John J. Harada and a National Science and Engineering Research Council Discovery Grant to Mark F. Belmonte.

REFERENCES

- Agati, G., Brunetti, C., Di Ferdinando, M., Ferrini, F., Pollastri, S., and Tattini, M. (2013). Functional roles of flavonoids in photoprotection: new evidence, lessons from the past. *Plant Physiol. Biochem.* 72, 35–45. doi: 10.1016/j.plaphy.2013.03.014
- Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657. doi: 10.1126/science.1086391
- Autran, D., Baroux, C., Raissig, M. T., Lenormand, T., Wittig, M., Grob, S., et al. (2011). Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* 145, 707–719. doi: 10.1016/j.cell.2011.04.014
- Baroux, C., Autran, D., Raissig, M. T., Grimanelli, D., and Grossniklaus, U. (2013). Parental contributions to the transcriptome of early plant embryos. *Curr. Opin. Genet. Dev.* 23, 72–74. doi: 10.1016/j.gde.2013.01.006
- Beeckman, T., De Rycke, R., Viana, R., and Inze, D. (2000). Histological study of seed coat development in *Arabidopsis thaliana*. *J. Plant Res.* 113, 139–148. doi: 10.1007/PL00013924
- Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. *Proc. Natl. Acad. Sci. U.S.A.* 110, E435–E444. doi: 10.1073/pnas.1222061110
- Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* 41, 7429–7437. doi: 10.1093/nar/gkt520
- Braybrook, S. A., and Harada, J. J. (2008). LECs go crazy in embryo development. *Trends Plant Sci.* 13, 624–630. doi: 10.1016/j.tplants.2008.09.008
- Brown, R. C., Lemmon, B. E., and Nguyen, H. (2003). Events during the first four rounds of mitosis establish three developmental domains in the syncytial endosperm of *Arabidopsis thaliana*. *Protoplasma* 222, 167–174. doi: 10.1007/s00709-003-0010-x
- Brown, R. C., Lemmon, B. E., Nguyen, H., and Olsen, O. A. (1999). Development of endosperm in *Arabidopsis thaliana*. *Sex. Plant Reprod.* 12, 32–42. doi: 10.1007/s004970050169
- Casson, S., Spencer, M., Walker, K., and Lindsey, K. (2005). Laser capture microdissection for the analysis of gene expression during embryogenesis of *Arabidopsis*. *Plant J.* 42, 111–123. doi: 10.1111/j.1365-313X.2005.02355.x
- Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 156, 1479–1491. doi: 10.1016/j.cell.2013.12.001
- Chen, H., Wang, F.-W., Dong, Y.-Y., Wang, N., Sun, Y.-P., Li, X.-Y., et al. (2012). Sequence mining and transcript profiling to explore differentially expressed genes associated with lipid biosynthesis during soybean seed development. *BMC Plant Biol.* 12:122. doi: 10.1186/1471-2229-12-122
- Chen, X. (2012). Small RNAs in development – insights from plants. *Curr. Opin. Genet. Dev.* 22, 361–367. doi: 10.1016/j.gde.2012.04.004
- Christian, M., Qi, Y., Zhang, Y., and Voytas, D. F. (2013). Targeted mutagenesis of *Arabidopsis thaliana* using engineered TAL effector nucleases. *G3 (Bethesda)* 3, 1697–1705. doi: 10.1534/g3.113.007104
- Cleland, J. (2013). World population growth: past, present and future. *Environ. Resour. Econ.* 55, 543–554. doi: 10.1007/s10640-013-9675-9676
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Costa, L. M., Yuan, J., Rouster, J., Paul, W., Dickinson, H., and Gutierrez-Marcos, J. F. (2012). Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr. Biol.* 22, 160–165. doi: 10.1016/j.cub.2011.11.059
- Curtin, S. J., Zhang, F., Sander, J. D., Haun, W. J., Starker, C., Baltes, N. J., et al. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol.* 156, 466–473. doi: 10.1104/pp.111.172981
- Day, R. C., Herridge, R. P., Ambrose, B. A., and Macknight, R. C. (2008). Transcriptome analysis of proliferating *Arabidopsis* endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant Physiol.* 148, 1964–1984. doi: 10.1104/pp.108.1.28108
- Day, R. C., McNoe, L., and Macknight, R. C. (2007). Evaluation of global RNA amplification and its use for high-throughput transcript analysis of laser-microdissected endosperm. *Int. J. Plant Genomics* 2007:61028. doi: 10.1155/2007/61028
- de Pater, S., Pinas, J. E., Hooikaas, P. J. J., and van der Zaal, B. J. (2013). ZFN-mediated gene targeting of the *Arabidopsis* protoporphyrinogen oxidase gene through *Agrobacterium*-mediated floral dip transformation. *Plant Biotechnol. J.* 11, 510–515. doi: 10.1111/pbi.12040
- Dean, G., Cao, Y., Xiang, D., Provart, N. J., Ramsay, L., Ahad, A., et al. (2011). Analysis of gene expression patterns during seed coat development in *Arabidopsis*. *Mol. Plant* 4, 1074–1091. doi: 10.1093/mp/ssr040
- Dean Rider, S., Henderson, J. T., Jerome, R. E., Edenberg, H. J., Romero-Severson, J., and Ogas, J. (2003). Coordinate repression of regulators of embryonic identity by PICKLE during germination in *Arabidopsis*. *Plant J.* 35, 33–43. doi: 10.1046/j.1365-313X.2003.01783.x
- Debeaujon, I., Léon-Kloosterziel, K. M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiol.* 122, 403–414. doi: 10.1104/pp.122.2.403
- Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M., et al. (2003). Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development. *Plant Cell* 15, 2514–2531. doi: 10.1105/tpc.014043.1
- Dekkers, B. J. W., Pearce, S., van Bolderen-Veldkamp, R. P., Marshall, A., Widera, P., Gilbert, J., et al. (2013). Transcriptional dynamics of two seed compartments with opposing roles in *Arabidopsis* seed germination. *Plant Physiol.* 163, 205–215. doi: 10.1104/pp.113.223511
- Farnham, P. J. (2009). Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* 10, 605–616. doi: 10.1038/nrg2636
- Feeeny, M., Frigerio, L., Cui, Y., and Menassa, R. (2013). Following vegetative to embryonic cellular changes in leaves of *Arabidopsis* overexpressing LEAFY COTYLEDON2. *Plant Physiol.* 162, 1881–1896. doi: 10.1104/pp.113.2.20996
- Feuillet, C., Leach, J. E., Rogers, J., Schnable, P. S., and Eversole, K. (2011). Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* 16, 77–88. doi: 10.1016/j.tplants.2010.10.005
- Finch-Savage, W. E., Cadman, C. S. C., Toorop, P. E., Lynn, J. R., and Hilhorst, H. W. M. (2007). Seed dormancy release in *Arabidopsis* Cvi by dry after-ripening, low temperature, nitrate and light shows common quantitative patterns of gene expression directed by environmentally specific sensing. *Plant J.* 51, 60–78. doi: 10.1111/j.1365-313X.2007.03118.x
- Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-120137
- Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342. doi: 10.1038/nature10452
- Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi: 10.1016/j.tibtech.2013.04.004
- Garcia, D., Fitz Gerald, J. N., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell* 17, 52–60. doi: 10.1105/tpc.104.027136
- Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jurgens, G., and Berger, F. (2003). *Arabidopsis* haiku mutants reveal new controls of seed size by endosperm. *Plant Physiol.* 131, 1661–1670. doi: 10.1104/pp.102.018762
- Gehring, M., Bubb, K. L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324, 1447–1451. doi: 10.1126/science.1171609
- Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE* 6:e23687. doi: 10.1371/journal.pone.0023687
- Giraudat, J., Hauge, B. M., Valon, C., Smalle, J., Parcy, F., and Goodman, H. M. (1992). Isolation of the *Arabidopsis* ABI3 gene by positional cloning. *Plant Cell* 4, 1251–1261. doi: 10.1105/tpc.4.10.1251

- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383
- Goldberg, R. B., de Paiva, G., and Yadegari, R. (1994). Plant embryogenesis: zygote to seed. *Science* 266, 605–614. doi: 10.1126/science.266.5185.605
- Gutierrez, L., Van Wuytswinkel, O., Castelain, M., and Bellini, C. (2007). Combined networks regulating seed maturation. *Trends Plant Sci.* 12, 294–300. doi: 10.1016/j.tplants.2007.06.003
- Haig, D., and Westoby, M. (1989). Parent-specific gene-expression and the triplid endosperm. *Am. Nat.* 134, 147–155. doi: 10.1086/284971
- Harada, J. J. (1997). “Seed maturation and control of germination,” in *Advances in Cellular and Molecular Biology of Plants*, Vol. 4, *Cellular and Molecular Biology of Seed Development*, eds B. A. Larkins and I. K. Vasil (Dordrecht: Springer), 545–592.
- Haughn, G., and Chaudhury, A. (2005). Genetic analysis of seed coat development in *Arabidopsis*. *Trends Plant Sci.* 10, 472–477. doi: 10.1016/j.tplants.2005.08.005
- Holdsworth, M. J., Bentsink, L., and Soppe, W. J. (2008). Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *New Phytol.* 179, 33–54. doi: 10.1111/j.1469-8137.2008.02437.x
- Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., et al. (2009). Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324, 1451–1454. doi: 10.1126/science.1172417
- Hsieh, T.-F. F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., et al. (2011). Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1755–1762. doi: 10.1073/pnas.1019273108
- Hsu, S.-D., Tseng, Y.-T., Shrestha, S., Lin, Y.-L., Khaleel, A., Chou, C.-H., et al. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.* 42, D78–D85. doi: 10.1093/nar/gkt1266
- Huang, D., Koh, C., Feurtado, J. A., Tsang, E. W. T., and Cutler, A. J. (2013). MicroRNAs and their putative targets in *Brassica napus* seed maturation. *BMC Genomics* 14:140. doi: 10.1186/1471-2164-14-140
- Huang, J., DeBowles, D., Esfandiari, E., Dean, G., Carpita, N. C., and Haughn, G. W. (2011). The *Arabidopsis* transcription factor LUH/MUM1 is required for extrusion of seed coat mucilage. *Plant Physiol.* 156, 491–502. doi: 10.1104/pp.111.172023
- Hu, J. H., Mitchum, M. G., Barnaby, N., Ayele, B. T., Ogawa, M., Nam, E., et al. (2008). Potential sites of bioactive gibberellin production during reproductive growth in *Arabidopsis*. *Plant Cell* 20, 320–336. doi: 10.1105/tpc.107.057752
- Ishida, T., Hattori, S., Sano, R., Inoue, K., Shirano, Y., Hayashi, H., et al. (2007). *Arabidopsis* TRANSPARENT TESTA GLABRA2 is directly regulated by R2R3 MYB transcription factors and is involved in regulation of GLABRA2 transcription in epidermal differentiation. *Plant Cell* 19, 2531–2543. doi: 10.1105/tpc.107.052274
- Jiang, J., Shao, Y., Li, A., Lu, C., Zhang, Y., and Wang, Y. (2013a). Phenolic composition analysis and gene expression in developing seeds of yellow- and black-seeded *Brassica napus*. *J. Integr. Plant Biol.* 55, 537–551. doi: 10.1111/jipb.12039
- Jiang, J., Shao, Y., Li, A., Zhang, Y., Wei, C., and Wang, Y. (2013b). FT-IR and NMR study of seed coat dissected from different colored progenies of *Brassica napus*–*Sinapis alba* hybrids. *J. Sci. Food Agric.* 93, 1898–1902. doi: 10.1002/jsfa.5986
- Jiang, Y., and Deyholos, M. K. (2010). Transcriptome analysis of secondary-wall-enriched seed coat tissues of canola (*Brassica napus* L.). *Plant Cell Rep.* 29, 327–342. doi: 10.1007/s00299-010-0824-x
- Johnson, C. S., Kolevski, B., and Smyth, D. R. (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell* 14, 1359–1375. doi: 10.1105/tpc.001404.covered
- Jones, S. I., Gonzalez, D. O., and Vodkin, L. O. (2010). Flux of transcript patterns during soybean seed development. *BMC Genomics* 11:136. doi: 10.1186/1471-2164-11-136
- Junker, A., Hartmann, A., Schreiber, F., and Baumlein, H. (2010). An engineer's view on regulation of seed development. *Trends Plant Sci.* 15, 303–307. doi: 10.1016/j.tplants.2010.03.005
- Kagaya, Y., Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H., et al. (2005a). Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in *Arabidopsis*. *Plant Cell Physiol.* 46, 300–311. doi: 10.1093/pcp/pci031
- Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., and Hattori, T. (2005b). LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46, 399–406. doi: 10.1093/pcp/pci048
- Kang, I.-H. H., Steffen, J. G., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* 20, 635–647. doi: 10.1105/tpc.107.055137
- Kawashima, T., and Goldberg, R. B. (2010). The suspensor: not just suspending the embryo. *Trends Plant Sci.* 15, 23–30. doi: 10.1016/j.tplants.2009.11.002
- Keith, K., Kraml, M., Dengler, N. G., and McCourt, P. (1994). fusca3: a heterochronic mutation affecting late embryo development in *Arabidopsis*. *Plant Cell* 6, 589–600. doi: 10.1105/tpc.6.5.589
- Khan, D., Millar, J. L., Girard, I. J., and Belmonte, M. F. (2014). Transcriptional circuitry underlying seed coat development in *Arabidopsis*. *Plant Sci.* 219–220, 51–60. doi: 10.1016/j.plantsci.2014.01.004
- Kohler, C., Wolff, P., and Spillane, C. (2012). Epigenetic mechanisms underlying genomic imprinting in plants. *Annu. Rev. Plant Biol.* 63, 331–352. doi: 10.1146/annurev-arplant-042811-105514
- Koornneef, M., Reuling, G., and Karssen, C. M. (1984). The isolation and characterization of abscisic-acid insensitive mutants of *Arabidopsis thaliana*. *Physiol. Plant.* 61, 377–383. doi: 10.1111/j.1399-3054.1984.tb06343.x
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Langridge, P., and Fleury, D. (2011). Making the most of “omics” for crop breeding. *Trends Biotechnol.* 29, 33–40. doi: 10.1016/j.tibtech.2010.09.006
- Lara, P., Onate-Sanchez, L., Abraham, Z., Ferrandiz, C., Diaz, L., Carbonero, P., et al. (2003). Synergistic activation of seed storage protein gene expression in *Arabidopsis* by ABI3 and two bZIPs related to OPAQUE2. *J. Biol. Chem.* 278, 21003–21011. doi: 10.1074/jbc.M210538200
- Lau, S., Slane, D., Herud, O., Kong, J., and Jurgens, G. (2012). Early embryogenesis in flowering plants: setting up the basic body pattern. *Annu. Rev. Plant Biol.* 63, 483–506. doi: 10.1146/annurev-arplant-042811-105507
- Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719
- Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8063–8070. doi: 10.1073/pnas.1003530107
- Le, B. H., Wagmaister, J. A., Kawashima, T., Bui, A. Q., Harada, J. J., and Goldberg, R. B. (2007). Using genomics to study legume seed development. *Plant Physiol.* 144, 562–574. doi: 10.1104/pp.107.100362
- Lefebvre, V., North, H., Frey, A., Sotta, B., Seo, M., Okamoto, M., et al. (2006). Functional analysis of *Arabidopsis* NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J.* 45, 309–319. doi: 10.1111/j.1365-313X.2005.02622.x
- Li, J., and Berger, F. (2012). Endosperm: food for humankind and fodder for scientific discoveries. *New Phytol.* 195, 290–305. doi: 10.1111/j.1469-8137.2012.04182.x
- Lipsa, F. D., Snowdon, R., and Friedt, W. (2011). Quantitative genetic analysis of condensed tannins in oilseed rape meal. *Euphytica* 184, 195–205. doi: 10.1007/s10681-011-0546-543
- Liu, H., Jin, T., Liao, R., Wan, L., Xu, B., Zhou, S., et al. (2012). miRFANs: an integrated database for *Arabidopsis thaliana* microRNA function annotations. *BMC Plant Biol.* 12:68. doi: 10.1186/1471-2229-12-68
- Liu, L., Qu, C., Wittkop, B., Yi, B., Xiao, Y., He, Y., et al. (2013a). A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. *PLoS ONE* 8:e83052. doi: 10.1371/journal.pone.0083052
- Liu, X., Lu, Y., Yuan, Y., Liu, S., Guan, C., Chen, S., et al. (2013b). De novo transcriptome of *Brassica juncea* seed coat and identification of genes for the biosynthesis of flavonoids. *PLoS ONE* 8:e71110. doi: 10.1371/journal.pone.0071110
- Lotan, T., Ohto, M., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., et al. (1998). *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93, 1195–1205. doi: 10.1016/S0092-8674(00)81463-4

- Lu, J., Zhang, C., Baulcombe, D. C., and Chen, Z. J. (2012). Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5529–5534. doi: 10.1073/pnas.1203094109
- Luerssen, H., Kirik, V., Herrmann, P., and Misera, S. (1998). FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in *Arabidopsis thaliana*. *Plant J.* 15, 755–764. doi: 10.1046/j.1365-313X.1998.00259.x
- Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., and Chaudhury, A. (2005). MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17531–17536. doi: 10.1073/pnas.0508418102
- Macquet, A., Ralet, M.-C., Kronenberger, J., Marion-Poll, A., and North, H. M. (2007). In situ, chemical and macromolecular study of the composition of *Arabidopsis thaliana* seed coat mucilage. *Plant Cell Physiol.* 48, 984–999. doi: 10.1093/pcp/pcm068
- Meinke, D., Muralla, R., Sweeney, C., and Dickerman, A. (2008). Identifying essential genes in *Arabidopsis thaliana*. *Trends Plant Sci.* 13, 483–491. doi: 10.1016/j.tplants.2008.06.003
- Meinke, D. W. (1992). A homeotic mutant of *Arabidopsis thaliana* with leafy cotyledons. *Science* 258, 1647–1650. doi: 10.1126/science.258.5088.1647
- Meinke, D. W., Franzmann, L. H., Nickle, T. C., and Yeung, E. C. (1994). Leafy cotyledon mutants of *Arabidopsis*. *Plant Cell* 6, 1049–1064. doi: 10.1105/tpc.6.8.1049
- Melkus, G., Rolletschek, H., Radchuk, R., Fuchs, J., Rutten, T., Wobus, U., et al. (2009). The metabolic role of the legume endosperm: a noninvasive imaging study. *Plant Physiol.* 151, 1139–1154. doi: 10.1104/pp.109.143974
- Miyawaki, K., Matsumoto-Kitano, M., and Kakimoto, T. (2004). Expression of cytokinin biosynthetic isopentenyltransferase genes in *Arabidopsis*: tissue specificity and regulation by auxin, cytokinin, and nitrate. *Plant J.* 37, 128–138. doi: 10.1046/j.1365-313X.2003.01945.x
- Monke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hahnel, U., et al. (2012). Toward the identification and regulation of the *Arabidopsis thaliana* ABI3 regulon. *Nucleic Acids Res.* 40, 8240–8254. doi: 10.1093/nar/gks594
- Mosher, R. A., Melnyk, C. W., Kelly, K. A., Dunn, R. M., Studholme, D. J., and Baulcombe, D. C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* 460, 283–286. doi: 10.1038/nature08084
- Mu, J., Tan, H., Zheng, Q., Fu, F., Liang, Y., Zhang, J., et al. (2008). LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*. *Plant Physiol.* 148, 1042–1054. doi: 10.1104/pp.108.126342
- Nakabayashi, K., Okamoto, M., Koshihara, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* 41, 697–709. doi: 10.1111/j.1365-313X.2005.02337.x
- Nakamura, S., Lynch, T. J., and Finkelstein, R. R. (2001). Physical interactions between ABA response loci of *Arabidopsis*. *Plant J.* 26, 627–635. doi: 10.1046/j.1365-313X.2001.01069.x
- Nelson, T., Tausta, S. L., Gandotra, N., and Liu, T. (2006). Laser microdissection of plant tissue: what you see is what you get. *Annu. Rev. Plant Biol.* 57, 181–201. doi: 10.1146/annurev.arplant.56.032604.144138
- Nesi, N., Debeaujon, I., Jond, C., Stewart, A. J., Jenkins, G. I., Caboche, M., et al. (2002). The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* 14, 2463–2479. doi: 10.1105/tpc.004127
- Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* 13, 2099–2114.
- Nishimura, N., Yoshida, T., Kitahata, N., Asami, T., Shinozaki, K., and Hirayama, T. (2007). ABA-Hypersensitive Germination1 encodes a protein phosphatase 2C, an essential component of abscisic acid signaling in *Arabidopsis* seed. *Plant J.* 50, 935–949. doi: 10.1111/j.1365-313X.2007.03107.x
- Nodine, M. D., and Bartel, D. P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.* 24, 2678–2692. doi: 10.1101/gad.1986710
- Nodine, M. D., and Bartel, D. P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* 482, 94–97. doi: 10.1038/nature10756
- Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y., and Yamaguchi, S. (2003). Gibberellin biosynthesis and response during *Arabidopsis* seed germination. *Plant Cell* 15, 1591–1604. doi: 10.1105/tpc.011650.ble
- Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J. J., et al. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell* 11, 407–416. doi: 10.1105/tpc.11.3.407
- Ohto, M. A., Floyd, S. K., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (2009). Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*. *Sex. Plant Reprod.* 22, 277–289. doi: 10.1007/s00497-009-0116-111
- Olsen, O. A. (2004). Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* 16, 214–228. doi: 10.1105/tpc.017111
- Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., and Giraudat, J. (1994). Regulation of gene expression programs during *Arabidopsis* seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell* 6, 1567–1582. doi: 10.1105/tpc.6.11.1567
- Penfield, S., Li, Y., Gilday, A. D., Graham, S., and Graham, I. A. (2006). *Arabidopsis* ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. *Plant Cell* 18, 1887–1899. doi: 10.1105/tpc.106.041277.1
- Penfield, S., Meissner, R. C., Shoue, D. A., Carpita, N. C., and Bevan, M. W. (2001). MYB61 is required for mucilage deposition and extrusion in the *Arabidopsis* seed coat. *Plant Cell* 13, 2777–2791. doi: 10.1105/tpc.010265.1
- Pourcel, L., Irani, N. G., Koo, A. J. K., Bohorquez-Restrepo, A., Howe, G. A., and Grotewold, E. (2013). A chemical complementation approach reveals genes and interactions of flavonoids with other pathways. *Plant J.* 74, 383–397. doi: 10.1111/tpl.12129
- Preston, J., Tatsumatsu, K., Kanno, Y., Hobo, T., Kimura, M., Jikumaru, Y., et al. (2009). Temporal expression patterns of hormone metabolism genes during imbibition of *Arabidopsis thaliana* seeds: a comparative study on dormant and non-dormant accessions. *Plant Cell Physiol.* 50, 1786–1800. doi: 10.1093/pcp/pcp121
- Qu, C., Fu, F., Lu, K., Zhang, K., Wang, R., Xu, X., et al. (2013). Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded *Brassica napus*. *J. Exp. Bot.* 64, 2885–2898. doi: 10.1093/jxb/ert148
- Raissig, M. T., Bemer, M., Baroux, C., and Grossniklaus, U. (2013). Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2. *PLoS Genet.* 9:e1003862. doi: 10.1371/journal.pgen.1003862
- Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* 8:e66428. doi: 10.1371/journal.pone.0066428
- Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M., and Lepiniec, L. (2005). LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Lett.* 579, 4666–4670. doi: 10.1016/j.febslet.2005.07.037
- Schneeberger, K., and Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci.* 16, 282–288. doi: 10.1016/j.tplants.2011.02.006
- Scott, R. J., Spielman, M., Bailey, J., and Dickinson, H. G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125, 3329–3341.
- Severin, A. J., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160
- Shamimuzzaman, M., and Vodkin, L. (2012). Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. *BMC Genomics* 13:310. doi: 10.1186/1471-2164-13-310
- Simbaya, J., Slominski, B. A., Rakow, G., Campbell, L. D., Downey, R. K., and Bell, J. M. (1995). Quality characteristics of yellow-seeded *Brassica* seed meals: protein, carbohydrates, and dietary fiber components. *J. Agric. Food Chem.* 43, 2062–2066. doi: 10.1021/jf00056a020
- Song, Q.-X. X., Liu, Y.-F. F., Hu, X.-Y. Y., Zhang, W.-K. K., Ma, B., Chen, S.-Y. Y., et al. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol.* 11:5. doi: 10.1186/1471-2229-11-5
- Sreenivasulu, N., and Wobus, U. (2013). Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annu. Rev. Plant Biol.* 64, 189–217. doi: 10.1146/annurev-arplant-050312-120215

- Stone, S. L., Braybrook, S. A., Paula, S. L., Kwong, L. W., Meuser, J., Pelletier, J., et al. (2008). *Arabidopsis* LEAFY COTYLEDON2 induces maturation traits and auxin activity: implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3151–3156. doi: 10.1073/pnas.0712364105
- Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., et al. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11806–11811. doi: 10.1073/pnas.201413498
- Suzuki, M., and McCarty, D. R. (2008). Functional symmetry of the B3 network controlling seed development. *Curr. Opin. Plant Biol.* 11, 548–553. doi: 10.1016/j.pbi.2008.06.015
- Tadros, W., and Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* 136, 3033–3042. doi: 10.1242/dev.033183
- Tang, X., Bian, S., Tang, M., Lu, Q., Li, S., Liu, X., et al. (2012). MicroRNA-mediated repression of the seed maturation program during vegetative development in *Arabidopsis*. *PLoS Genet.* 8:e1003091. doi: 10.1371/journal.pgen.1003091
- Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700
- Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108
- To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., et al. (2006). A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. *Plant Cell* 18, 1642–1651. doi: 10.1105/tpc.105.039925
- Troncoso-Ponce, M., Kilaru, A., Cao, X., Durrett, T. P., Fan, J., Jensen, J. K., et al. (2011). Comparative deep transcriptional profiling of four developing oilseeds. *Plant J.* 68, 1014–1027. doi: 10.1111/j.1365-313X.2011.04751.x
- Verdier, J., Dessaint, F., Schneider, C., Abirached-darmency, M., and Bitonti, B. (2013). A combined histology and transcriptome analysis unravels novel questions on *Medicago truncatula* seed coat. *J. Exp. Bot.* 64, 459–470. doi: 10.1093/jxb/err313
- Vicente-Carbajosa, J., and Carbonero, P. (2005). Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int. J. Dev. Biol.* 49, 645–651. doi: 10.1387/ijdb.052046jc
- Wang, F., and Perry, S. E. (2013). Identification of direct targets of FUSCA3, a key regulator of *Arabidopsis* seed development. *Plant Physiol.* 161, 1251–1264. doi: 10.1104/pp.112.212282
- Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., and Kohler, C. (2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet.* 6:e1001152. doi: 10.1371/journal.pgen.1001152
- West, M. A. L., Yee, K. M., Danao, J., Zimmerman, J. L., Fischer, R. L., Goldberg, R. B., et al. (1994). LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. *Plant Cell* 6, 1731–1745. doi: 10.1105/tpc.6.12.1731
- Western, T. L., Burn, J., Tan, W. L., Skinner, D. J., Martin-mccaffrey, L., Moffatt, B. A., et al. (2001). Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*. *Plant Physiol.* 127, 998–1011. doi: 10.1104/pp.010410.upon
- Western, T. L., Skinner, D. J., and Haughn, G. W. (2000). Differentiation of mucilage secretory cells of the *Arabidopsis* seed coat. *Plant Physiol.* 122, 345–356. doi: 10.1104/pp.122.2.345
- Willmann, M. R., Mehalick, A. J., Packer, R. L., and Jenik, P. D. (2011). MicroRNAs regulate the timing of embryo maturation in *Arabidopsis*. *Plant Physiol.* 155, 1871–1884. doi: 10.1104/pp.110.171355
- Windsor, J. B., Symonds, V. V., Mendenhall, J., and Lloyd, A. M. (2000). *Arabidopsis* seed coat development: morphological differentiation of the outer integument. *Plant J.* 22, 483–493. doi: 10.1046/j.1365-313x.2000.00756.x
- Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T., et al. (2011). High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet.* 7:e1002126. doi: 10.1371/journal.pgen.1002126
- Xiang, D., Venglat, P., Tibiche, C., Yang, H., Risseuw, E., Cao, Y., et al. (2011). Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. *Plant Physiol.* 156, 346–356. doi: 10.1104/pp.110.171702
- Xie, K., and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* 6, 1975–1983. doi: 10.1093/mp/sst119
- Young, R. E., McFarlane, H. E., Hahn, M. G., Western, T. L., Haughn, G. W., and Samuels, L. (2008). Analysis of the Golgi apparatus in *Arabidopsis* seed coat cells during polarized secretion of pectin-rich mucilage. *Plant Cell* 20, 1623–1638. doi: 10.1105/tpc.108.058842
- Yu, C.-Y. (2013). Molecular mechanism of manipulating seed coat coloration in oilseed *Brassica* species. *J. Appl. Genet.* 54, 135–145. doi: 10.1007/s13353-012-0132-y
- Zhang, F., Maeder, M. L., Unger-wallace, E., Hoshaw, J. P., Reyon, D., Christian, M., et al. (2010). High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12028–12033. doi: 10.1073/pnas.0914991107
- Zhang, K., Lu, K., Qu, C., Liang, Y., Wang, R., Chai, Y., et al. (2013). Gene silencing of BnTT10 family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. *PLoS ONE* 8:e61247. doi: 10.1371/journal.pone.0061247
- Zhao, Y.-T., Wang, M., Fu, S.-X., Yang, W.-C., Qi, C.-K., and Wang, X.-J. (2012). Small RNA profiling in two *Brassica napus* cultivars identifies microRNAs with oil production- and development-related expression and new small RNA classes. *Plant Physiol.* 158, 813–823. doi: 10.1104/pp.111.187666

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 July 2014; accepted: 26 August 2014; published online: 12 September 2014.
Citation: Becker MG, Hsu S-W, Harada JJ and Belmonte MF (2014) Genomic dissection of the seed. *Front. Plant Sci.* 5:464. doi: 10.3389/fpls.2014.00464

This article was submitted to *Plant Evolution and Development*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2014 Becker, Hsu, Harada and Belmonte. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.