


Integration of omics approaches to understand oil/protein content during seed development in oilseed crops

Manju Gupta¹  · Pudota B. Bhaskar¹ · Shreedharan Sriram¹ · Po-Hao Wang¹

Received: 4 April 2016 / Accepted: 11 October 2016 / Published online: 27 October 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Oilseed crops, especially soybean (*Glycine max*) and canola/rapeseed (*Brassica napus*), produce seeds that are rich in both proteins and oils and that are major sources of energy and nutrition worldwide. Most of the nutritional content in the seed is accumulated in the embryo during the seed filling stages of seed development. Understanding the metabolic pathways that are active during seed filling and how they are regulated are essential prerequisites to crop improvement. In this review, we summarize various omics studies of soybean and canola/rapeseed during seed filling, with emphasis on oil and protein traits, to gain a systems-level understanding of seed development. Currently, most (80–85%) of the soybean and rapeseed reference genomes have been sequenced (950 and 850 megabases, respectively). Parallel to these efforts, extensive omics datasets from different seed filling stages have become available. Transcriptome and proteome studies have detected preponderance of starch metabolism and glycolysis enzymes to be the possible cause of higher oil in *B. napus* compared to other crops. Small RNAome studies performed during the seed filling stages have revealed miRNA-mediated regulation of transcription factors, with the suggestion that this interaction could be responsible for transitioning the seeds from embryogenesis to maturation. In addition, progress made in dissecting the regulation of de novo fatty

acid synthesis and protein storage pathways is described. Advances in high-throughput omics and comprehensive tissue-specific analyses make this an exciting time to attempt knowledge-driven investigation of complex regulatory pathways.

Keywords Seed development · Seed composition · Omics · Systems biology · Soybean · Rapeseed/canola

Introduction

Seed development is a dynamic process coordinated by the three distinct organs of a seed: embryo, endosperm and seed coat (Goldberg et al. 1994; Miller et al. 1999; Moïse et al. 2005; Weber et al. 2005; Lafon-Placette and Köhler 2014; Radchuk and Borisjuk 2014). Seed developmental programs, governed by plant genetics and environmental conditions, culminate in determination of seed composition. Seed development occurs in a precise and orchestrated manner after two fertilization steps trigger the formation of a seed (Le et al. 2007). The diploid embryo and triploid endosperm represent the filial lineages, whereas the testa represents the diploid maternal generation. The endosperm initially develops as a syncytium (Brown et al. 2003) then cellularizes to encircle the developing embryo. The endosperm either remains to feed the growing embryo and to store starch, lipids and/or storage proteins as reserves in mature seeds, such as in corn or is fully absorbed, except for the aleurone layer, by the developing embryo such as in soybean (Sreenivasulu and Wobus 2013). The seed coat, which develops by differentiation of two ovule integuments, not only protects the developing embryo and endosperm but also has other roles such as metabolic control of seed development and dormancy, metabolism of

Communicated by N. Sreenivasulu.

Pudota B. Bhaskar, Shreedharan Sriram and Po-Hao Wang contributed equally to the review.

✉ Manju Gupta
mgupta@dow.com

¹ Dow AgroSciences, 9330 Zionsville Road, Indianapolis, IN 46268, USA

nutrients from parent plant, and disease resistance (Islam et al. 2003; Weber et al. 2005). Each organ is programmed to perform its function individually and by interactions with the other two organs (Lorenz et al. 2014; Radchuk and Borisjuk 2014).

Oilseed crops are grown for the oils contained in their seed and/or fruits and include soybean (*Glycine max*), canola/rapeseed (*Brassica napus*), sunflower, peanuts, palm, flax (linseed), and castor bean. Seed oils are used for various applications, including food, feed, industrial oils, and cosmetics (Thelen and Ohlrogge 2002). In addition to oil, protein is the major source of nutrition in oilseed crops. The oil and protein content in seeds varies depending on the species. For example, the grain of wheat (*Triticum aestivum*) contains 1–2% oil and 8–15% protein (Shewry 2009), while soybean seed contains 20% oil and 40% protein (Hill and Breidenbach 1974; Ohlrogge and Kuo 1984) and canola/rapeseed (*B. napus*) seed contains 40% oil and 15% protein (Norton and Harris 1975). Canola is an improved cultivar of rapeseed created through breeding that contains <2% erucic acid and <30 micromoles of glucosinolates per gram of seed and has multiple health benefits (Lin et al. 2013).

Oilseed crops store fatty acid as triacylglycerols (TAGs) in the oil bodies (OBs) within embryo cytoplasm. TAG synthesis takes place in the plastids through the coordination of multiple metabolic pathways including Calvin cycle, carbon assimilation (glycolysis), starch metabolism and oxidative pentose phosphate pathway (OPPP). Sugars are transported from the vegetative tissue to the endosperm (Lorenz et al. 2014) and then to the embryo (source to sink tissue), and it is assimilated through a parallel glycolytic pathway operational in the cytosol as well as in plastid. Carbon compounds resulting from glycolysis, e.g., phosphoenolpyruvate (PEP), are transported from the cytosol to the plastids via PEP-phosphate translocators, in exchange for inorganic phosphate (reviewed in Fischer and Weber 2002), for the synthesis of acetyl-CoenzymeA (acetyl-CoA). The latter serves as a substrate in the first step of TAG synthesis (reviewed in Voelker and Kinney 2001; Schultz and Ohlrogge 2002). Following three key steps of fatty acid synthesis (FAS), acyl chain elongation, desaturation and termination, the resulting free fatty acids are exported to cytoplasm where they become esterified to coenzyme A (CoA) and serve as substrates in the final key step of TAG synthesis operational in the endoplasmic reticulum (ER; reviewed in Voelker and Kinney 2001; Schultz and Ohlrogge 2002). Redox equivalents (NADPH) for FAS are supplied by the OPPP functional in plastids. TAGs are exported to cytoplasm and stored in OBs between two polar lipid (PL) monolayers. The OBs are prominent in oleosins and it prevents OBs from coalescence in concert with PL (Katavic et al. 2006).

In addition to oil, seed storage proteins (SSPs) are major sources of reduced carbon and nitrogen that are critical to the life cycle of plants. Soybean seed contains two major storage proteins, namely β -conglycinin (7S; Harada et al. 1989) and glycinin (11S; Nielsen et al. 1989). Together they constitute 70% of total seed protein (Meinke et al. 1981). There are 5 versions of glycinin (G1–G5) and 3 versions of β -conglycinin (α , α' and β) proteins. *B. napus* contains two major SSPs, 2S napin (or 2S albumin) and 12S cruciferin (Höglund et al. 1992). They constitute 20 and 60%, respectively, of the total mature seed protein. All soybean and *B. napus* SSPs are initially synthesized on rough endoplasmic reticulum and then transported via the Golgi apparatus to the protein storage vesicles (PSV) for storage in the differentiated cells of embryo (reviewed in Chrispeels 1991; Herman and Larkins 1999). As stated earlier, this process occurs only during the cell expansion phase of the embryogenesis and during the seed-filling stages (Rubel et al. 1972; Meinke et al. 1981; Le et al. 2007). The accumulated SSPs and TAGs survive seed desiccation and then mobilized through degradation to supply carbon and nitrogen to the growing seedling.

Dicot seed development has been reviewed, as understood through one or more of the omics approaches focused on systems biology (Le et al. 2007; Gallardo et al. 2008; Thompson et al. 2009; Sreenivasulu and Wobus 2013). The goal of this review is to highlight recent advances in the omics-level understanding of de novo FAS, fatty acid (FA) accumulation and protein storage during seed development in soybean (*Glycine max*) and rapeseed/canola (*Brassica napus*). Studies on miRNA during seed development have been included in this review, since miRNAs regulate developmental and physiological processes and play an essential role in seed development (He and Hannon 2004). This recent progress and the availability of large global datasets reveal the pathways toward genetic improvement of oilseed crops.

Development of genomics resources

The availability and accessibility of reference genome sequences for crops of economic and agronomic importance are driving accelerated crop improvement. Following the release of the *Arabidopsis* reference genome in 2000 (Arabidopsis Genome Initiative 2000), reference-quality genome sequences have been established for several important dicot crop species (Table 1). As part of the process, new methods were developed to tackle the complexities involved with crop genomes, such as very large size, highly repetitive content and polyploidy (Michael and VanBuren 2015). The massive throughput and cost-effectiveness of next generation sequencing (NGS)

Table 1 Description of various dicot plant genomes sequenced and the technologies used for sequencing

Name	Scientific name	Year	Type	Genome size ^a	Technologies used ^b	PMID
Arabidopsis	<i>Arabidopsis thaliana</i>	2000	Model	125	S	11130711
Grape	<i>Vitis vinifera</i>	2007	Crop	475	S	17721507
Lotus	<i>Lotus japonicus</i>	2008	Model	472	S	18511435
Castor bean	<i>Ricinus communis</i>	2010	Crop	320	S	20729833
Soybean	<i>Glycine max</i> *	2010	Crop	1115	S	20075913
Chinese cabbage	<i>Brassica rapa</i>	2011	Crop	485	I	21873998
Cocoa	<i>Theobroma cacao</i>	2011	Crop	430	S, R, I	21186351
Pigeon pea	<i>Cajanus cajan</i>	2011	Crop	833	S, I	22057054
Medicago	<i>Medicago truncatula</i>	2011	Model	454	S, R, I	22089132
Potato	<i>Solanum tuberosum</i>	2011	Crop	844	S, R, I	21743474
Flax	<i>Linum usitatissimum</i>	2012	Crop	373	I	22757964
Tomato	<i>Solanum lycopersicum</i>	2012	Crop	900	S, R, I	22660326
Chickpea	<i>Cicer arietinum</i>	2013	Crop	738	S, I	23354103
Brassica	<i>Brassica oleracea</i>	2014	Crop	630	I, R	24852848
Canola	<i>Brassica napus</i> **	2014	Crop	1130	S, R, I	25146293
Coffee	<i>Coffea canephora</i>	2014	Crop	710	S, R, I	25190796
Hot pepper	<i>Capsicum annuum</i>	2014	Crop	3480	I	24441736

* Cultivar W82

** Cultivar Darmor-*bzh*^a In megabases (Mb)^b S Sanger, R 454 (Roche), I Illumina

technologies, coupled with better experimental and computational approaches, have made it possible for comparative genomics to deliver new insights into genomic variations (within and across species), trait mapping and phylogenomics studies (Morrell et al. 2011). Reference genome assemblies facilitate the identification of functional genes and pathways that drive seed development programs and that regulate molecular processes affecting seed size and storage (Sreenivasulu and Wobus 2013).

Soybean is the first legume species to have its genome sequenced, in 2010. The cultivar Williams 82 was sequenced and assembled using an approach that integrated whole genome shotgun-based Sanger sequencing data with physical and genetic maps to create the reference sequence. The genome assembly comprises ~950 megabases (Mb), representing about 85% of the predicted 1115-Mb genome (Schmutz et al. 2010). The latest version of the genome assembly, available at Phytozome (<https://phytozome.jgi.doe.gov/>; Goodstein et al. 2012), has been annotated with ~56,050 protein-coding genes, of which ~1500 genes are annotated as homologs of *Arabidopsis* genes with potential roles in lipid metabolism (Ye et al. 2014). Availability of these genomic resources, coupled with existing genetic and QTL information, such as the 183 seed oil QTLs annotated as part of Soybase (Grant et al. 2010), have been foundational to understanding soybean seed

development and related processes and has contributed to establishing soybean as a model crop legume.

Brassica napus is not only an important crop but also a model for allopolyploid evolution. Leveraging the past efforts to establish genome sequences for *B. rapa* (Wang et al. 2011) and *B. oleracea* (Liu et al. 2014), the *B. napus* genome (winter double low oilseed rape cultivar Darmor-*bzh*) was sequenced and assembled using an integrated approach combining 454, Sanger and Illumina sequencing technologies. The reference sequence (final assembly size of ~850 Mb) and its annotated ~101,000 gene models were mined for factors that could improve breeding efforts for improved seed oil quality, lipid composition and pathogen resistance (Chalhoub et al. 2014). In particular, ~2300 homologs of lipid biosynthesis genes were identified within the genome, many of which are conserved with the progenitor A and C subgenomes of *B. rapa* and *B. oleracea*, respectively. These genes include many of the known genes related to oil biosynthesis and regulation.

Assembled reference genomes provide a foundation for investigating features that link genotype and phenotype. For this purpose, the combination of genome re-sequencing using different cultivars, high-throughput genotyping and comparative genomics offers significant advantages in enabling new genetic mapping strategies (Feuillet et al. 2011) and genome-wide association studies (GWAS)

(Korte and Farlow 2013). In soybean, high-throughput Illumina re-sequencing of wild (*G. soja*), landrace and cultivated (*G. max*) lines was utilized to perform a GWAS, which revealed high linkage disequilibrium in the soybean genomes (Lam et al. 2010; Zhou et al. 2015). The investigations also identified several selective signals corresponding to domestication and improvement traits, 96 of which overlapped locations of well-known oil content QTLs and were found to contain 21 fatty acid biosynthesis genes. A GWAS performed specifically on 175 soybean lines with information on oil content revealed 6 loci that had strong GWAS signals, 5 of which were located within previously identified oil content QTLs. These signals are prime genomic locations in which to identify genes with functional relevance to oil content (Lam et al. 2010; Zhou et al. 2015). In addition, data from the USDA Germplasm Resources Information Network database (<http://www.ars-grin.gov/>) have been utilized to compare germplasm with normal or high seed protein content. GWAS analysis discovered significant associations with seed protein concentrations for 40 SNPs located across 10 different chromosomes and significant associations with seed oil content for 25 SNPs across 12 different chromosomes (Hwang et al. 2014). In *B. napus*, a recent study utilized a GWAS approach on an association panel containing 89 genotypes using ~4000 SNP markers to identify 17 significant associations for seed glucosinolate content and 5 associations for seed hemicellulose content, many in loci contained in previously reported QTLs (Gajardo et al. 2015). In another study, genome-wide genotype information and phenotype data for 12 seed quality traits (including oil content, oleic acid concentration, and linoleic acid concentration) were analyzed using a GWAS approach to identify 112 significantly associated seed quality SNPs (Korber et al. 2016).

Transcriptome changes during seed development

The genomics resources discussed above lay the foundation for studies that aim to understand the role and dynamics of functional elements, such as genes and their expression, within the genome. Advances in transcriptomics, primarily hybridization- and sequencing-based technologies, have resulted in numerous large gene expression data sets. The popularity of RNA-Seq has been increasing, as NGS technology costs continue to decrease (Lister et al. 2008; Garg and Jain 2013). In addition to cost, RNA-Seq has technical advantages, such as profiling a broader dynamic range of expression, increased specificity and sensitivity, and the ability to detect alternative splicing events (Marioni et al. 2008; Fu et al. 2009; Trapnell et al. 2010), particularly when prior genome or

transcriptome reference information is unavailable (Xia et al. 2011; Wu et al. 2015).

The availability of the soybean reference genome along with improvements in global transcriptome profiling have provided a powerful way to study expression changes at the genome-wide scale across multiple stages of soybean seed development. High-throughput transcriptome sequencing using Illumina technology was performed on different tissue types and developmental stages to establish an expression atlas accessible to the research community (Libault et al. 2010; Severin et al. 2010). These datasets identified tissue- and stage-specific expression patterns related to seed development. For instance, expression levels of 27,945 genes in developing seeds 28 days after flowering (DAF) were increased as compared to the levels in flowers (Severin et al. 2010). This is the largest number of differentially expressed genes observed between any of the seed or tissue developmental stages. The results suggested important roles for sucrose transport, SSPs (oleosin, lectin and beta-conglycinin), lipid biosynthesis (lipoxygenase) and seed coat development (seed coat BURP domain) during the seed development process. Gene expression clustering and gene ontology (GO) enrichment analyses indicated an over-representation of GO terms related to nutrient reservoir activity in the late seed development stages and pointed to multiple highly expressed genes that are involved in the seed filling process with functions related to those described above. These findings have been corroborated in subsequent studies. RNA-Seq-based expression patterns indicated that a large number of histone-encoding genes and proline-rich protein encoding genes showed high expression during early seed developmental stages, indicating rapid cell growth and development and initiating the differentiation processes that give rise to the embryo, seed coat, cotyledons and endosperm. During later developmental stages, such as 100–200 and 400–500 mg cotyledon weights, genes encoding storage proteins, such as beta-conglycinin, oleosins, glycinin, and lipoxygenase are highly expressed and point to activation of nutrient accumulation and storage and oxidation of polyunsaturated fatty acids, consistent with biological processes known to be active in those stages as well as earlier studies in *Arabidopsis* (Peng and Weselake 2011). During seed desiccation and the final stages of seed development, highly expressed genes annotated as hydrophilic proteins and late embryogenesis abundant proteins were over-represented, indicative of low water conditions in plants. (Jones and Vodkin 2013; O'Rourke et al. 2014).

Lipid biosynthesis-related genes exhibit programmed expression during seed development. To study this aspect, transcriptome data across multiple soybean seed developmental stages were evaluated to identify expression patterns of lipid biosynthesis-related genes, revealing that

genes such as *FAD2-2B* and *FAD2-2C* were highly expressed in early stages of seed formation and the *FAD2-1B* and *FAD2-1A* were highly expressed at later seed stages (O'Rourke et al. 2014). To unravel differences in seed oil content and composition among multiple soybean genotypes, transcriptomes of soybean seeds from the mid-maturation stage were sequenced, showing that SSPs (such as glycinin and beta-conglycinin), lipid metabolism genes (such as lipoxygenases and *FAD2-1B*) and oil body proteins (such as oleosins) were among the most abundantly expressed seed transcripts. Comparison of gene expression patterns with transcript polymorphisms detected across the different genotypes provided evidence that differential expression of genes related to lipid biosynthesis is potentially responsible for the different fatty acid compositions seen in these genotypes (Goettel et al. 2014b). This is exemplified in soybean lines having higher stearic acid levels, which were found to have no *FAD2-1A* activity and a non-synonymous mutation in *FAB2C*. The large amount of data from these and other transcriptome profiling studies (Chaudhary et al. 2015) have been made available at public repositories such as Gene Expression Omnibus (GEO) and NCBI Sequence Read Archive to facilitate further investigations.

Omics studies in *B. napus* exploring oil-related biological processes are enabling an understanding of factors that could impact genetic improvement of oilseed rape. In one *B. napus* study, global gene expression profiles from a number of seed developmental stages with differing fatty acid metabolism were generated using cDNA chip hybridization (across >8000 ESTs cloned from seed) and helped to identify functional pathways of fatty acid biosynthesis and regulation, which were conserved in relation to *Arabidopsis* (Niu et al. 2009). A crucial role of starch metabolism was observed early in seed development of *B. napus*, which could be contributing to higher oil content in *B. napus* than in *Arabidopsis*. In the same study, nearly 30 starch metabolism-related and OPPP genes were identified for differential expression in *B. napus* as compared to that in *Arabidopsis*. It was suggested that hormone (auxin and jasmonate) signaling exhibited important roles in FA metabolism. In another study, candidate genes involved in oil-related biological processes in *B. napus* cultivars were investigated using RNA-Seq on pod tissues across different developmental stages (5–7, 15–17 and 25–27 DAF). The genes were found to be involved in carbohydrate metabolism, amino acid metabolism and lipid metabolism. Subsequent analyses using previously reported QTLs of oil content and orthologous lipid-related genes from *Arabidopsis* showed that a total of 33 differentially expressed genes that were potentially involved in lipid metabolism overlapped with known QTL locations, marking them as good candidates for further investigation (Xu et al. 2015).

In summary, advances in resolution, throughput and cost of transcriptomic data generation and interpretation have led to major insights into the atlas of expression pattern changes across the seed developmental stages in soybean and rapeseed/canola. Overall the new data do corroborate with known pathways discovered previously by other techniques. The integration of these datasets to create a comprehensive view across multiple studies would be beneficial and improve the ability to mine these datasets for additional knowledge. For example, in soybean, a number of different transcriptome datasets have been integrated to create a compendium of information stored within the Soybean Functional Genomics Database (SFGD) (<http://bioinformatics.cau.edu.cn/SFGD/>), which has functionalities that allow interrogating gene expression patterns across soybean genome, including ~1500 genes involved in multiple acyl-lipid metabolism pathways (Yu et al. 2014).

Proteomics of developing seeds

Proteomics studies have been conducted to follow protein expression during seed development in oilseed crops. Studies of the 5 seed-filling stages of rapeseed (*B. napus* 'Reston'; Hajduch et al. 2006) and soybean (*Glycine max* 'Maverick'; Hajduch et al. 2005; Agrawal et al. 2008) identified hundreds of non-redundant (NR) proteins differentially expressed during seed filling by utilizing multiple high-throughput protein separation technologies. Representative proteins were mapped onto metabolic pathways, e.g., carbon assimilation, to understand fatty acid metabolism. Proteins from metabolism, destination and storage, and energy were prevalent. The prevalence of these classes revealed high demand on the developing embryos for metabolites. Proteomics of high and low oil isogenic varieties in sunflower, another oilseed crop, identified 77 differentially expressed proteins (DEPs) revealing a tight link between oil content and carbohydrate metabolism/protein synthesis (Hajduch et al. 2007).

During seed development, sets of proteins representing polysaccharide, amino acid, fatty acid and SSP metabolisms displayed dynamic yet consistent (over replicates) expression trends in rapeseed and soybean (Hajduch et al. 2006; Agrawal et al. 2008). Accordingly, fatty acids and lipids proteins peaked at 4 weeks after flowering (WAF) in both rapeseed and soybean and then declined. Whereas storage proteins peaked at 4 WAF in both the crops but continued increasing through 6 WAF (Agrawal et al. 2008). Similar expression trends were also observed in transcriptome analysis (Niu et al. 2009; Li et al. 2015), and reflected regulatory control during seed development.

Another notable aspect of the proteomes was the presence of multiple isoforms per protein (MI) or multiple

isoforms per protein within a family of proteins (MFI). For example, over 38% of the soybean proteins that are classified as seed storage, seed maturation and sucrose binding were prevalent in MFIs, with 2–30 isoforms detected. Storage proteins were especially prevalent in MFIs: 8 NR proteins produced 148 MFIs, consisting of 7–30 isoforms per protein. These proteins were comprised of glycinin (G1, G2 and other), β -conglycinin (α , α' and β) and agglutinin in A complex (Harada et al. 1989; Nielsen et al. 1989). Similarly, in rapeseed, 33% of the identified NR proteins contained more than 2 MFIs; of these, six of the proteins were the cruciferins (Hajduch et al. 2006). These MFIs could either result from post-transcriptional alternative splicing (reviewed in Smith et al. 1989; Godovac-Zimmermann et al. 2005) or from post-translational modifications, including phosphorylation (reviewed in Kersten et al. 2006) or glycosylation (Lei and Reeck 1987), or a combination of both. These protein architectures reveal various dimensions of protein regulation that could be important during seed development. Overall, the seed filling stages presented 119 NR SSPs in soybean, compared to 71 observed in rapeseed (Agrawal et al. 2008).

Protein phosphorylation is a ubiquitous and dynamic post-translational modification that controls diverse biological processes, including in plants (Huber and Hardin 2004; Mukherji 2005; Pawson and Scott 2005). Phosphoproteomes were analyzed from developing seeds of rapeseed and soybean to establish quantitative expression profiles and to locate phosphorylation sites in the proteins (Agrawal and Thelen 2006; Meyer et al. 2012). The phosphoproteins (PPs) represented the major classes of seed development proteins reported earlier, e.g., energy, metabolism, protein destination and signal transduction (Hajduch et al. 2006; Agrawal et al. 2008). PPs were detected in the enzymes of the FAS pathway (e.g., stearoyl-acyl-carrier-protein desaturase and beta-ke-toacyl-ACP synthase I) as well as in SSPs (cruciferin MFIs) of rapeseed (Agrawal and Thelen 2006). Stearoyl-acyl-carrier-protein desaturase I has been implicated in the regulation of cell growth and development and the defense/stress responses (Shanklin and Cahoon 1998; Kachroo et al. 2001).

One of the PPs, a 14-3-3, was prominently detected in the developing seeds of *B. napus* and soybean (Agrawal and Thelen 2006; Agrawal et al. 2008). This protein is ubiquitous in eukaryotes and known to regulate metabolic pathways by interacting with the components of transcription factor complexes, and in turn activating them (Aitken 2002; Fulgosi et al. 2002). To prove the involvement of 14-3-3 in *Arabidopsis* seed development, it was shown that two isoforms of 14-3-3 could selectively bind proteins of metabolic pathways, including that of carbon assimilation and de novo FAS (Swatek et al. 2011). Further studies are needed to understand the biological implications of 14-3-3 in *Arabidopsis* as well as in oilseed crops.

Highly abundant proteins of the soybean proteomes described above, e.g., β -conglycinin, glycinin, agglutinin (lectin), SBP and LOX, were also abundant in the soybean transcriptome (see the Transcriptome section; Severin et al. 2010). Other interesting results came from a comparative proteomic study. LOX and SBP were abundant in soybean seed but underrepresented or absent in *B. napus* (Hajduch et al. 2006; Agrawal et al. 2008). The soybean transcriptome atlas confirmed the identification of 72 LOX genes in the genome, of which only 3 are highly and significantly expressed during seed development. LOX acts on polyunsaturated fatty acid to form polyunsaturated fatty acid hydroperoxides; the latter can be converted into aldehydes and alcohols, which would impart lower flavor quality in soybean (Narvel et al. 1998; reviewed in Porta and Rocha-Sosa 2002). SBP might have a role in sucrose translocation-dependent physiological processes (Ripp et al. 1988). High expression of these proteins in soybean but not *B. napus* point to differential functional roles in seed development, and warrant further study.

Proteomics of tissue subtypes contained within a developing seed could provide more direct information regarding metabolic architecture and pathway interactions. Proteomics of endosperm and embryo isolated from the developing seeds of *B. napus* at multiple time points (Lorenz et al. 2014) showed that endosperm contains the entire set of central metabolic pathways, which makes it a self-sustained and metabolically competent tissue. A high number of transcription factors and regulatory components were also identified in the endosperm, indicating its role in signaling for coordinated seed development. Differences were observed in the metabolic capability of endosperm versus embryo at 15 DAF and 20 DAF, based on the spot volume. For example, carbohydrate metabolism was high in endosperm at both the stages; amino acid metabolism was higher in embryo at 15 DAF but higher in endosperm at 20 DAF; energy metabolism was higher in endosperm at 15 DAF but higher in embryo at 20 DAF (embryos became capable of photosynthesis at this stage); lipid metabolism was higher in embryos in both the stages. Magnetic resonance-based studies have shown endosperm to be a major location for lipid deposition in the early stages of seed development (Borisjuk et al. 2013). The results show pre-determined and precise spatiotemporal expression patterns of FA and storage proteins, which indicate complex hierarchical regulation at transcriptional, post-transcriptional and post-translational levels during seed development.

Proteomics of developing and mature seed reveal distinguishing features of higher oil in *B. napus*

A comparative proteomics study of rapeseed (oil-rich) and soybean (protein-rich) developing seeds revealed up to

threefold higher expression of proteins in the FAS pathway in rapeseed as compared to soybean (Hajduch et al. 2006; Agrawal et al. 2008). Enzymes of the carbon assimilation and glycolysis pathways were higher in number (~48%) and abundance (~80%) in rapeseed compared to soybean. Niu et al. (2009) reported higher copy number of genes of the above pathways in *B. napus* (Huyou 15) transcriptomes than in *Arabidopsis*. One of the genes encoding the acyl carrier protein 1 of the FAS pathway was present in 16 copies as compared to 3-ketoacyl-ACP synthase that was present in only 2 copies. The results pointed toward transcriptional regulation of FAS genes. Taken together, the observation suggested increased flow of carbon from sucrose to glycolysis and eventually to de novo FAS in *B. napus*. This significant difference could be responsible for higher oil in rapeseed compared to soybean (Agrawal et al. 2008; Niu et al. 2009). In addition, Rubisco large subunit was observed to be substantially high in the developing seeds of rapeseed (Agrawal et al. 2008). There were 10 MIs of Rubisco in rapeseed, contributing to 4% of the total enzyme at the peak, but only one MI of Rubisco in soybean was observed. It has been shown that Rubisco can work without the Calvin cycle to assimilate the CO₂ released by the plastidial pyruvate dehydrogenase complex to maximize the efficiency of FAS in embryo (Schwender et al. 2004). High abundance of Rubisco (Hajduch et al. 2006; Demartini et al. 2011; Gan et al. 2013) supported more efficient recycling of CO₂ by Rubisco resulting in higher de novo FAS in rapeseed as compared to soybean. *B. napus* transcriptome was also observed to express both Rubisco small unit and Ribulose-biphosphate carboxylase large subunit at high level during majority of seed development (Niu et al. 2009). These studies supported the idea that plastids are functionally programmed for high oil production in rapeseed due to coordination of cytosolic and plastidial carbon assimilation and efficient de novo FAS (Demartini et al. 2011).

In addition to the seed filling stages, proteomics of mature seeds has highlighted additional features of higher oil traits. The proteomes of mature seeds of two winter-type *B. napus* lines, low (36.49%; LO) and high (55.19%; HO) in oil content, were compared to understand differences in protein composition and oil body characteristics (Gan et al. 2013). The results showed that late embryogenesis proteins (LEA; Dong et al. 2004) were differentially expressed in HO and LO lines and might be correlated with the number and size of the OBs. These observations were corroborated by transmission electron microscopy and confocal scanning microscopy studies, which showed that the HO line contained smaller cells with thicker cell walls, more densely packed OBs, and sparse protein storage vacuoles (PSV), whereas the LO line contained larger cells with thinner walls, fewer OBs, and more PSVs. In other

words, total protein content and oil content were inversely correlated in the *B. napus* HO and LO lines. Nevertheless, no enzymes involved in de novo FA synthesis were detected in mature seeds of either the LO or HO *B. napus* lines, indicating that accumulation of FAs is over prior to the maturation stage.

In summary, proteomic studies of the oilseed crops have contributed remarkable advances to our understanding of seed development. Proteomics technology has been developed to fractionate, identify and quantify proteins and to map those proteins onto metabolic pathways to understand fine regulation of genes during seed development. Development of proteomic databases and associated software has allowed sharing of the information in the scientific community.

Metabolomics of seed development

Metabolomics is a specialized form of analytical biochemistry that has been used to assist in the biochemical analysis of complex mixtures. It is considered a robust, sensitive, and powerful technology (Nakabayashi and Saito 2013). Studies on seed-based metabolomics in dicot crops are still limited (Chaudhary et al. 2015). Metabolomics requires powerful and sensitive technologies for analysis of small-molecule metabolites to reveal perturbations, whether environmental or genetic, in metabolic compositions (Clarke et al. 2013). Analytical techniques, such as Mass Spectrometry (MS) and/or Nuclear Magnetic Resonance spectroscopy (NMR), coupled with metabolite databases have been used in previous studies to determine the number of metabolites, to detect new metabolites, metabolite composition, and natural metabolite variations, and to even compare metabolite profiles for biosafety assessment (Oikawa et al. 2008; Lin et al. 2014; Harrigan et al. 2015; Kortensniemi et al. 2015; Chebrolu et al. 2016). Various separation and detection techniques have been used for metabolomics studies, such as Gas Chromatography-Mass Spectroscopy (GC-MS), Nuclear Magnetic Resonance (NMR), Liquid Chromatography-MS (LC-MS), Fourier Transform Ion Cyclotron Resonance-MS (FTICT-MS), a combination of GC-MS, Ultra-Performance Liquid Chromatography-Tandem MS (UPLC-MS/MS), and Capillary Electrophoresis-MS (CE-MS) (Oikawa et al. 2006; Clarke et al. 2013; Lin et al. 2014; Tan et al. 2015). GC-MS has been proven to be the most efficient, sensitive and a reliable tool for conducting various metabolomics studies in dicot crops. Recently, publications have demonstrated that NMR has several benefits as a metabolite profiling technique, including ease of sample preparation, high sample throughput, identification of a broad range of chemical compounds in a single experiment, and generation of

quantitative data (Harrigan et al. 2015; Han et al. 2016). Sometimes a combination of one or more analytical platforms has been used in various metabolomics studies.

Tan et al. (2015) investigated the metabolome of developing canola seeds to understand the dynamic metabolic changes that occur during oil accumulation in seeds. A total of 443 putative metabolites were identified in the oil accumulation stage in the seeds and in the silique walls of canola following 3 treatments (leaf detachment, phloem peeling and silique darkening). The metabolite profiles showed that high concentrations of metabolites in seeds, when transferred from silique walls, induced a subset of genes related to FA synthesis and sugar metabolism to increase the metabolic flux, eventually leading to enhancing seed-oil content (Tan et al. 2015). In another study, Kortessniemi et al. investigated the compositions of ripened and developing seed of *B. napus* and *B. rapa* using metabolomics (Kortessniemi et al. 2015). Higher levels of polyunsaturated fatty acids, especially alpha-linolenic acid and sucrose, were more characteristic to *B. rapa* than *B. napus*, while sinapine levels were higher in *B. napus* (Kortessniemi et al. 2015). Metabolomics approaches in canola could also be applied to study metabolomic diversity between seeds with altered fatty acid composition (transgenic-based or mutagenesis-derived) and wild-type canola. Recently, a rapid analytical method based on ¹H-NMR technology was developed to differentiate high and low erucic acid rapeseed *Brassicac*s (Han et al. 2016). Erucic acid (C22:1) is an undesirable fatty acid in canola seed, and levels must be less than 5% in commodity canola oil. This NMR-based method can effectively distinguish low-erucic acid rapeseed (LEAR) from high-erucic-acid rapeseed oil with convenient sample pretreatment and data acquisition by hierarchical cluster analysis (Han et al. 2016).

Lin et al. (2014) characterized soybean seed metabolic profiles from 29 different soybean cultivars, which led to identification of 169 named metabolites and 104 seed metabolites that significantly varied in their levels across all tested cultivars. In addition, metabolite–metabolite correlations and interactions were studied, resulting in the construction of a seed metabolic network map based on all 169 metabolites identified (Lin et al. 2014). This might be helpful for potential metabolic engineering and molecular breeding efforts aiming to enhance seed quality in soybean. Harrigan et al. demonstrated the application of ¹H-NMR for assessment of seed metabolic diversity among soybean varieties differing in yield potential (Harrigan et al. 2015). A total of 27 diverse metabolites were quantified among the 9 varieties used in this study and confirmed that metabolite variability is influenced by both selective breeding and environment. The techniques and methods used in both of these studies could be applied to advance soybean metabolomics in developing seed.

Advances in database development and bioinformatics tools are still lagging behind for the field of seed metabolomics. Databases for integrating, mining and visualizing various types of omics and phenotypic data from soybeans have been developed (<http://soymetDB.org>; <http://soykb.org>; Joshi et al. 2010, 2012). The potential findings derived from various metabolic studies performed on different soybean tissues will provide a basis for the improvement of soybean seed quality and perhaps even yield.

Omics integration for improved understanding of seed development

Transcriptomic, metabolomic and metabolic flux data were generated and integrated to uncover the metabolic program in developing soybean seed (*G. max* ‘Evans’) and to determine mature seed composition (Li et al. 2015). In the metabolomic analysis across 5 stages of seed development, encompassing 25–50 DAF, about 400 analyte peaks were identified. Of these, 273 metabolites were above the detection limit; 148 of these were chemically identified, while 125 remained unidentified. For transcriptomic analysis, levels of 37,593 transcripts represented by probes on the *Glycine max* Affymetrix chip were analyzed; 2879 of these were differentially expressed over seed development (q -value = 0.01). The Plant/Eukaryotic and Microbial Metabolomics Systems Resource (PMR) and MetNet systems biology tools were used for interactive analyses of all datasets. The amino acid asparagine was identified as the major form of nitrogen that is imported from the vegetative plant tissue during seed development. Further, a particular asparaginase enzyme was predicted to interconvert amino acids for protein synthesis during seed fill. Consistent with measured starch accumulation, the levels of transcripts involved in starch synthesis and related to transporters decreased after 45 DAF, while those of starch degrading enzyme increased after 45 DAF. The oil content increased steadily until 40 DAF and then declined. This pattern was matched with the transcripts involved in fatty acid metabolism, whereas genes involved in β -oxidation (involved in mobilizing the stored oil reserve) increased expression after 45 DAF. Statistical analysis of the transcriptomic and metabolomic data identified the pathways where the most variability occurred during seed development.

Early seed development had the highest metabolic flux through the gamma-aminobutyric acid (GABA) shunt, the tricarboxylic acid (TCA) cycle, and the OPPP through coordination of plastid, mitochondria and cytosol functions (Li et al. 2015). The transcriptomic and metabolomic data depicted similar changes. Overall, the study indicated consistent expression for the majority of the transcriptome,

except for a set of transcripts that varied in expression during seed development. On the other hand, the majority of metabolite levels changed during seed development. It was suggested that the smaller number of metabolites analyzed could be a reason for this observation. The other reason could be that expression differences are more amplified at the metabolite level than at transcription or translation, since metabolite levels account for the catalytic role of proteases following translation.

Collakova et al. (2013) reported metabolic and transcription analyses of soybean embryos from 10 different stages of seed development and integrated the data sets to gain insight into the metabolic processes. The transcriptomic analysis revealed the expression of 41,619 genes at one time point during seed development. In the early stages of embryo development, 55 polar metabolites, including sugars, sugar alcohols, sugar acids, amino acids, organic amines and alcohols, carboxylic acids and phenolic compounds, were detected using GC–MS and UPLC–FLD (Ultra-High Performance Liquid Chromatography). The metabolite analysis revealed a steady increase of fatty acids and proteins up to 25 days, when a plateau was reached for both. The levels of individual fatty acids started to decrease after 40 days, while the protein levels were maintained until the end of the time course. Two main transition phases for metabolic and transcriptional reprogramming were studied. The first occurs when dividing and differentiating cells transition into cell elongation during early seed filling, when the metabolism changes from heterotrophic to photoheterotrophic. The second occurs when the elongating cells at the seed filling stage turn on seed maturation and desiccation processes, to prepare seeds for dormancy and photoheterotrophic metabolism switches to heterotrophic. Clustering analysis revealed functionally related metabolites and transcripts operational in different developmental and metabolic programs. These clusters are a resource for further analysis that will likely identify potential targets for metabolic engineering.

The role of miRNA in seed development and the miRNA-mediated transcription factor network

Epigenetic regulation (including transcriptional and post-transcriptional) plays an important role in seed development, during which genomic imprinting or parent-of-origin gene expression is heavily involved in early endosperm development. The mechanism is extensively characterized in monocot and dicot plants (detailed reviews in Bai and Settles 2015; Kohler and Lafon-Placette 2015). In addition to the imprinting-mediated mechanism via DNA methylation, small RNAs (sRNAs) play important roles in gene

regulation during plant growth, development and reproduction (Jin et al. 2013; Borges and Martienssen 2015). MicroRNAs (miRNA), which are ~21-nt sRNAs, are mediators of gene expression regulation in plants via mRNA degradation or translational inhibition (Brodersen et al. 2008). The emerging evidence supports critical roles of miRNAs in multiple steps of seed development, seed germination and seedling growth. However, information regarding miRNA involvement in seed development in oilseed crops is limited, with only a few studies reported. Herein, we summarize studies on soybean and rapeseed seed-abundant miRNAs and genes targeted by miRNAs, and describe the significance of miRNA-mediated regulatory networks through transcription factors (TFs) in seed development.

miRNAs in soybean seed development

Song et al. (2011) conducted high-throughput sRNA and degradome (a method to analyze the pattern of mRNA degradation; Thomson et al. 2011) analyses from 15-DAF soybean (*G. max* ‘Heinong44’) seed and identified 55 annotated and 26 novel miRNAs. Through degradome sequencing, a total of 145 genes were identified as miRNA targets to annotated miRNAs and 25 as targets of new miRNAs (Song et al. 2011). The majority (82%) of miRNA targets were TFs, such as those from the ARF, GRF, NAC, MYB, and TCP families. Moreover, Shamimuzzaman and Vodkin (2012) conducted a soybean (*G. max* ‘Williams82’) degradome study using cotyledon and seed coat across the early to late seed maturation stages. Consistent with Song’s study, ARF, MYB, TCP, NF-Y, GRF, HD-ZIP, PPR, SBP and NAC TF family members were identified as major targets for miRNAs that were common in both cotyledon and seed coat tissues. Furthermore, multiple versions of SBP, MYB, ARF, NAC and HD-ZIPs TFs were identified in both tissue types as predicted targets for miR156, miR159, miR160, miR164 and miR166, respectively. Tissue-specific miRNA-regulated targets were also identified. For example, miR1513 and miR2109 (cotyledon) and miR393 and miR1523 (seed coat) were identified in different tissues, which implies a tissue-specific regulatory role for these particular miRNAs. Importantly, it was suggested that due to the large number of miRNA targets identified in the late seed maturation stage, miRNA-mediated regulation could be involved in shifting the developmental program from immature to maturation stage.

Zabala et al. (2012) reported a sRNAome study from soybean (*G. max* ‘Williams’, ‘Richland’, ‘PI194639’, and ‘1462312’) from multiple tissues, including whole seeds (12–14 DAF), immature seed coat and cotyledon. miRNAs with highly preferential expression in either seed or vegetative tissues were observed. For example, miR167 and

miR1512 were highly abundant in seed coats, whereas miR156 and miR3522 were highly abundant in cotyledons. Goettel et al. (2014a) conducted another sRNAome study to identify the miRNA population in soybean (*G. max* ‘Jack’) cotyledons at five seed developmental stages (S2, 3, 4, 6, and 8). This report showed that a number of miRNAs were highly expressed (e.g., miR156, miR166, miR319, miR159, miR164, miR167, miR482, miR1508, miR1510 and miR3522). In addition, the biological pathways of these miRNA targets were characterized. The cotyledon miRNA targets were predominantly involved in RNA metabolism and transcriptional regulation, with a large number of TFs such as ARF and TCP or those containing a CCAAT box.

miRNAs in *Brassica napus* seed development

An sRNAome study from developing (7–42 DAF) and mature (50 DAF) *B. napus* (PFB-2 from Embrapa) seeds identified highly expressed miRNAs (e.g., miR156, miR159, miR166, miR167 and miR824) (Korbes et al. 2012). The predicted targets of miRNAs abundant in mature seeds were mainly TFs involved in plant development, e.g., SPLs, ARF, NAC, SCL, and TOE, suggesting that these TFs may be more vital during early seed development. These results were consistent with results in soybean seeds.

In another extensive study in *B. napus* (DH12075) investigating the roles of miRNA in seed development, sRNAs were profiled from whole seeds at nine different seed developmental stages (10–50 DAF) and different tissues types, including radical, hypocotyls, cotyledon, embryo, endosperm and seed coat (Huang et al. 2013). Briefly, the miRNAs were much more abundant in embryo than in either endosperm or seed coat. The top six most abundant miRNA families expressed in seed were: miR156 (48% of total miRNAs reads), miR159, miR172, miR167, miR158 and miR166. The predicted targets of those highly abundant miRNAs were SPLs (miR156), MYBs (miR159), AP2-like TFs (miR172), ARFs (miR167), NAC (miR166) TFs, again implying the importance of these TF families in seed development.

Although different developmental stages and genetic backgrounds were used across the two species of soybean and rapeseed/canola, the findings of the miRNA and miRNA-target studies converged to reveal that: (1) a number of miRNAs were co-identified from most of the studies, such as miR156, miR159, miR160, miR164, miR166 and miR167. This observation highlights the crucial role of these miRNAs in regulating the downstream targets during seed development. (2) Through degradome sequencing and prediction of miRNA targets, TFs were the major miRNA targets. e.g., SPLs, MYBs, AP2 and ARFs

and NAC. Such miRNA targets emphasized the complexity of regulatory networks, via TFs, for seed development (described below). (3) The spatial expression patterns of particular miRNAs were identified from either cotyledons or seed coats, which highlighted the importance of tissue-specific miRNAs during seed development.

miRNA-mediated TF regulatory networks in seed development

The majority of miRNA targets identified during soybean and rapeseed seed development were TFs, including SPLs, MYBs, AP2 and ARFs and NAC. The roles of these TFs during distinct seed developmental stages are well characterized in *Arabidopsis* and rice (Fig. 1). In this section, we briefly summarize the role of these key TFs during *Arabidopsis* seed development and compare recent findings in different crops.

SPLs (miR156)-mediated network in early embryogenesis

In *Arabidopsis*, miR156 targets *SBP-like Protein10* (SPL10) and SPL11, both are involved in early morphogenesis of embryos (Nodine and Bartel 2010). In a loss-of-function *dcl1* line, which compromises the miRNA biosynthesis pathway, *SPL10* and *SPL11* transcripts were increased by ~150-fold, resulting in an abnormal embryo phenotype at the eight-cell stage of early embryo development. The disruption of miR156-mediated repression of SPL10/SPL11 also promotes precocious expression of multiple genes, indicating that the normal function of a miR156-SPLs regulatory module is required to repress the precocious accumulation of the transcripts normally expressed in maturation phase (Nodine and Bartel 2010). These results highlight the role of the miR156/SPLs regulatory module in negatively regulating phase transition and seed maturation.

ARFs (miR160/miR167)-mediated network in early embryogenesis

A loss-of-function miR160 mutant in *Arabidopsis*, *foc* (*floral organs in carpels*), exhibits impairment of embryo development resulting in the formation of aberrant seeds, suggesting a crucial role of miR160 during embryogenesis (Liu et al. 2010). Previous studies showed that Auxin Response Factors (ARFs) are negatively regulated by miR160 (Mallory et al. 2005; Wang et al. 2005; Liu et al. 2007). ARFs are transcription factors involved in auxin signal transduction during many stages of plant growth development via regulation of auxin response genes. In *Arabidopsis*, the crucial roles of ARFs are well characterized in distinct biological events and specific tissues,

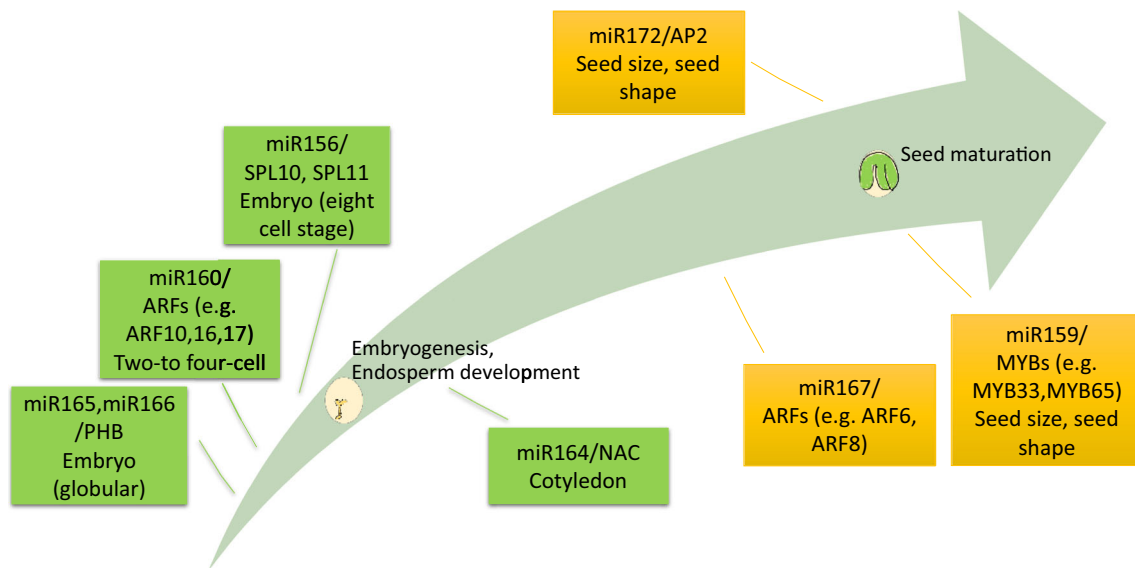


Fig. 1 Diagram illustrating the miRNA-mediated network within the developing seed. *Green* and *orange* boxes represent the miRNAs involved during early embryogenesis and late embryogenesis (seed maturation), respectively. *Each* box indicates a particular miRNA,

their target transcription factor genes, and when the abnormal phenotype was observed based on evidence corresponding to that in *Arabidopsis* or rice mutants

including seeds and embryos (Okushima et al. 2005; Wang et al. 2005; Weijers et al. 2006; Liu et al. 2010). The mutation of ARF5 impairs the initiation of the body axis in the early embryo (Hardtke and Berleth 1998), although ARF7 and ARF12 transcripts were detected throughout all embryo developmental stages (Hardtke et al. 2004). ARF10, ARF16 and ARF17 transcripts were highly increased in the miR160 *loc* mutant (Liu et al. 2010). Moreover, miR167, which targets ARF6 and ARF8, is preferentially expressed in rice seed, suggesting its involvement during seed development (Niu et al. 2009). These numerous studies together show that ARF TFs are crucial during seed development and that the correct patterning of the ARFs, as mediated by miRNAs, may ensure normal seed development.

NAC (miR164)-mediated network in early embryogenesis

miR164-mediated regulation of NO APICAL MERISTEM (NAC)-domain proteins plays a crucial role in embryo patterning and cotyledon development. Expression of miR164-resistant NAC (by introducing a NAC gene with mutations within the miR164-complementary site without changing the amino acids) results in embryonic development defects (i.e., cotyledon orientation defects), suggesting the importance of the proper miR164-mediated regulation of NAC for seed development (Mallory et al. 2004). Moreover, overexpression of miR164 leads to a cup-shaped cotyledon (Aida et al. 1997). miR164/NAC regulatory modules are key

regulatory components essential for normal seed embryogenesis and morphogenesis.

MYBs (miR159)-mediated network in late embryogenesis and maturation

During late embryogenesis, cell expansion and elongation play a key role and account for seed size. MYB TFs play crucial roles in broad biological functions (e.g., gene regulation, secondary metabolism, and environmental stress responses). miR159-mediated regulation of MYB TFs could be associated with seed size, since a double mutation of miR159 (miR159ab) causes small seeds with increased MYB33 and MYB65 expression (Allen et al. 2007). Additionally, MYBs play essential roles in seed coat development, since changes in AtMYB56, AtMYB5, and AtMYB61 expression patterns displayed effects on seed coat development (Penfield et al. 2001; Gonzalez et al. 2009; Zhang et al. 2013). For instance, MYB56 may serve as a positive regulator of seed size through regulating the genes that are involved in cell division and expansion, leading to the proper growth of cell number of the outer integument layer. The spatial and temporal expression of these MYBs is consistent with their function in seed development. Furthermore, genes in the R2R3-MYB superfamily have been identified in soybean (244 genes) and canola (76 genes) (Du et al. 2012; Chen et al. 2016). Since the function in seed development of most of the MYBs in oilseed crops still needs to be determined, further work should be done to characterize the MYBs in oilseed development.

AP2/ERF (miR172)-mediated network in late embryogenesis

The plant APETALA2 (AP2) and ethylene-responsive element (ERF) families are part of the AP2/ERF super family of TFs. The first AP2 protein was identified in *Arabidopsis* for its role in flower and seed development (Jofuku et al. 1994). AP2 TFs comprise one of the largest plant TF families and regulate various processes of growth, development, and stress responses. AP2 genes are regulated by miR172 in *Arabidopsis* (Aukerman and Sakai 2003). At least one AP2 TF plays a role in seed development and acts in all three major seed compartments: embryo, endosperm and seed coat. *Arabidopsis ap2* mutants displayed impairment of normal seed development (Ohto et al. 2005, 2009).

In summary, the above descriptions highlight how the TF-regulated networks present during early embryogenesis, endosperm development and/or seed maturation, and the miRNAs, as fine-tuners of the spatial and temporal expression of TFs, work together to ensure functional seed development. A recent study in maize (Li et al. 2016) also identified miRNAs involved in seed development, which corresponded with those previously identified in *Arabidopsis*, indicating that conserved pathways are present in monocot and dicot plants. Thus, the knowledge gained from the model species may be leveraged in the studies of oilseed crops to expedite the understanding of miRNA-mediated TF networks in seed development.

Future perspective

Genomics and transcriptomics technologies produce massive amounts of data that can broadly characterize DNA and RNA states across crop plants, such as genome-wide sequence variation across a population of genotypes or global differential gene expression between developmental stages. While experimental designs that utilize omics-based approaches generate a large number of hypotheses related to a biological trait of interest, challenges still exist in utilizing these approaches to pinpoint the key factors controlling a particular trait of interest. Hence, better experimental strategies at both small and large scales are required to more efficiently translate this vast array of information into specific traits in improved crops. These strategies could combine GWAS studies to map the trait to chromosomal locations with detailed analysis of the putative genes underlying those QTLs to identify specific gene(s) conferring the trait(s).

Additionally, advances in genomics technologies have helped to assemble reference genomes as a foundational resource. However, challenges exist in fully understanding

the dynamic nature of the functional elements within the genome. Next phases in understanding plant biology, such as aspects of seed development, at the systems level could employ an ENCODE (Encyclopedia of DNA elements; Consortium 2012) approach aimed at building a comprehensive look at how DNA coding and non-coding elements, epigenetic modifications, and transcriptome states vary and interact across developmental stages. Furthermore, foundational genome resources, composed of reference genomes assembled from individual varieties, should be upgraded to pan-genomes and pan-transcriptomes to more comprehensively represent traits present across diverse genotypes within crop species. Emerging sequencing technologies and library construction methods that are pushing the boundaries on throughput, read length, and long-range genomic information will further unlock the potential of these omics layers.

Whereas outstanding progress has been made in creating the proteomes of developing seeds in *B. napus* and soybean, there is opportunity to develop full proteomes of the same. Proteins expressed during the seed filling stages can have different physical, as well as functional, characteristics due to post-translational modifications. These characteristics may only be found through the use of different techniques for their extraction, analysis and quantitation.

Metabolic pathways for de novo oil biosynthesis are well established. Progress has been made to identify key enzymes or families that are higher in number and abundance in higher oil crops. However, regulation of these pathways must be understood for genetic engineering of the trait. Similarly, manipulation of SSPs requires better understanding of the function of each of the protein isoforms and the overall regulation of the families.

Currently, studies focused on seed-specific metabolites are limited in rapeseed/canola and soybean oilseed crops. Towards this end, better bioinformatics tools and integrated databases, supported by relevant software and computational analysis, are needed. These tools would aid in the development of both a comprehensive reference metabolome and a metabolite atlas for oilseed crop improvement.

Epigenetic regulation plays a key role in a variety of biological events, including seed development. Genomic imprinting and miRNA-mediated regulation orchestrates the complex gene network that leads to proper seed development. Although we did not focus on genomic imprinting in this review, the importance of imprinting has been demonstrated in both monocot (e.g., maize and rice) and dicot (e.g., *Arabidopsis*) plants. For instance, several genes have been identified that act maternally to regulate seed size, which highlights the potential of imprinting for crop seed improvement. However, little is known about imprinting in oilseed crops, creating opportunity to fully study the regulatory mechanisms of imprinting and seed size control.

Further studies in the oilseed crops will be needed to decipher the genetic pathway of genomic imprinting using the genetic framework built for other species.

The understanding of miRNA-mediated regulatory networks during seed development has remarkably progressed in *Arabidopsis* and rice with the rapid evolution of NGS technology, but little is known about the interplay of these different networks. For example, the loss-of-function *ap2* mutant in *Arabidopsis* increases seed size and enhances seed filling (Ohto et al. 2005), yet the detailed AP2-mediated pathway is unclear. Also, in addition to the TF-mediated network, epialleles (heritable epigenetic variants of genes) were identified that control seed size, which adds to the complexity of the seed regulatory network (Zhang et al. 2015). As extensive NGS datasets are being generated in different plant species and the capability to integrate the transcriptome, sRNA transcriptome, ChIP-seq and degradome data improves, there is an opportunity to further elucidate how the miRNA-mediated TFs regulate downstream genes during seed development and to identify the genetics behind novel seed traits (e.g., seed size, yield, oil content) in the oilseed crops.

Embryo, endosperm and seed coat interact in a hierarchical way to orchestrate seed development. Dissection of these biological processes with omics is needed to allow systems-level views of FA and SSP accumulation that will allow identification of master networks as targets for gene(s) manipulation for crop improvement. Furthermore, omics studies need to be combined with genetical approaches, e.g., gene mapping and GWAS, to understand gene function and regulation in the genomic context. Tools for this effort include zinc finger nucleases and CRISPR/Cas9 to induce targeted mutation(s) in one or more genes simultaneously (Zhang et al. 2014; Petolino 2015). Advances in high-throughput omics and computational analyses make this an exciting time to attempt knowledge-driven dissection of complex regulatory pathways.

Author contribution statement Each of the authors has contributed to one or more omic topics described in this review. BBP, SS and PW contributed equally to this review. MG is the corresponding author. The authors have no conflict of interest to claim.

Acknowledgements The authors would like to thank Rodrigo Sarria and Steve Rounsley for their support and advice, and Steve Rounsley and Cory Christensen for critical review of this manuscript.

References

Agrawal GK, Thelen JJ (2006) Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Mol Cell Proteomics* 5:2044–2059

- Agrawal GK, Hajduch M, Graham K, Thelen JJ (2008) In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. *Plant Physiol* 148:504–518
- Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997) Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. *Plant Cell* 9:841–857
- Aitken A (2002) Functional specificity in 14-3-3 isoform interactions through dimer formation and phosphorylation. Chromosome location of mammalian isoforms and variants. *Plant Mol Biol* 50:993–1010
- Allen RS, Li J, Stahle MI, Dubroué A, Gubler F, Millar AA (2007) Genetic analysis reveals functional redundancy and the major target genes of the *Arabidopsis* miR159 family. *Proc Natl Acad Sci USA* 104:16371–16376
- Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell* 15:2730–2741
- Bai F, Settles AM (2015) Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Front Plant Sci*. doi:10.3389/fpls.2014.00780
- Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* 16:727–741
- Borisjuk L, Neuberger T, Schwender J et al (2013) Seed architecture shapes embryo metabolism in oilseed rape. *Plant Cell* 25:1625–1640
- Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science* 320:1185–1190
- Brown R, Lemmon B, 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
- Chalhoub B, Denoeud F, Liu S et al (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950–953
- Chaudhary J, Patil GB, Sonah H, Deshmukh RK, Vuong TD, Valliyodan B, Nguyen HT (2015) Expanding omics resources for improvement of soybean seed composition traits. *Front Plant Sci*. doi:10.3389/fpls.2015.01021
- Chebrolov KK, Fritschi FB, Ye S, Krishnan HB, Smith JR, Gillman JD (2016) Impact of heat stress during seed development on soybean seed metabolome. *Metabolomics* 12:1–14
- Chen B, Niu F, Liu W-Z et al (2016) Identification, cloning and characterization of R2R3-MYB gene family in canola (*Brassica napus* L.) identify a novel member modulating ROS accumulation and hypersensitive-like cell death. *DNA Res* 23:101–114
- Chrispeels MJ (1991) Sorting of proteins in the secretory system. *Annu Rev Plant Biol* 42:21–53
- Clarke JD, Alexander DC, Ward DP, Ryals JA, Mitchell MW, Wulff JE, Guo L (2013) Assessment of genetically modified soybean in relation to natural variation in the soybean seed metabolome. *Sci Rep*. doi:10.1038/srep03082
- Collakova E, Aghamirzaie D, Fang Y et al (2013) Metabolic and transcriptional reprogramming in developing soybean (*Glycine max*) embryos. *Metabolites* 3:347–372
- Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74
- Demartini DR, Jain R, Agrawal G, Thelen JJ (2011) Proteomic comparison of plastids from developing embryos and leaves of *Brassica napus*. *J Proteome Res* 10:2226–2237

- Dong J, Keller WA, Yan W, Georges F (2004) Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. *Planta* 218:483–491
- Du H, Yang S-S, Liang Z, Feng B-R, Liu L, Huang Y-B, Tang Y-X (2012) Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol*. doi:10.1186/1471-2229-12-106
- Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K (2011) Crop genome sequencing: lessons and rationales. *Trends Plant Sci* 16:77–88
- Fischer K, Weber A (2002) Transport of carbon in non-green plastids. *Trends Plant Sci* 7:345–351
- Fu X, Fu N, Guo S et al (2009) Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genom*. doi:10.1186/1471-2164-10-161
- Fulgosi H, Soll J, de Faria Maraschin S, Korthout HA, Wang M, Testerink C (2002) 14-3-3 proteins and plant development. *Plant Mol Biol* 50:1019–1029
- Gajardo HA, Wittkop B, Soto-Cerda B et al (2015) Association mapping of seed quality traits in *Brassica napus* L. using GWAS and candidate QTL approaches. *Mol Breeding* 35:1–19
- Gallardo K, Thompson R, Burstin J (2008) Reserve accumulation in legume seeds. *C R Biol* 331:755–762
- Gan L, C-y Zhang, X-d Wang et al (2013) Proteomic and comparative genomic analysis of two *Brassica napus* lines differing in oil content. *J Proteome Res* 12:4965–4978
- Garg R, Jain M (2013) RNA-Seq for transcriptome analysis in non-model plants. *Methods Mol Biol* 1069:43–58
- Godovac-Zimmermann J, Kleiner O, Brown LR, Drukier AK (2005) Perspectives in spicing up proteomics with splicing. *Proteomics* 5:699–709
- Goettel W, Liu ZR, Xia J, Zhang WX, Zhao PX, An YQ (2014a) Systems and evolutionary characterization of microRNAs and their underlying regulatory networks in soybean cotyledons. *PLoS One*. doi:10.1371/journal.pone.0086153
- Goettel W, Xia E, Upchurch R, Wang ML, Chen P, An YQ (2014b) Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content. *BMC Genom*. doi:10.1186/1471-2164-15-299
- Goldberg RB, De Paiva G, Yadegari R (1994) Plant embryogenesis: zygote to seed. *Science* 266:605–614
- Gonzalez A, Mendenhall J, Huo Y, Lloyd A (2009) TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. *Dev Biol* 325:412–421
- Goodstein DM, Shu S, Howson R et al (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40:D1178–D1186
- Grant D, Nelson RT, Cannon SB, Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res* 38:D843–D846
- Hajduch M, Ganapathy A, Stein JW, Thelen JJ (2005) A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiol* 137:1397–1419
- Hajduch M, Casteel JE, Hurrelmeyer KE, Song Z, Agrawal GK, Thelen JJ (2006) Proteomic analysis of seed filling in *Brassica napus*. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. *Plant Physiol* 141:32–46
- Hajduch M, Casteel JE, Tang S, Hearne LB, Knapp S, Thelen JJ (2007) Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. *J Proteome Res* 6:3232–3241
- Han J, Lu C, Li Y, Deng Z, Fu B, Geng Z (2016) Discrimination of rapeseeds (*Brassica napus* L.) based on the content of erucic acid by 1H NMR. *Eur Food Res Technol* 242:441–447
- Harada JJ, Barker SJ, Goldberg RB (1989) Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell* 1:415–425
- Hardtke CS, Berleth T (1998) The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J* 17:1405–1411
- Hardtke CS, Ckurshumova W, Vidaurre DP et al (2004) Overlapping and non-redundant functions of the Arabidopsis auxin response factors MONOPTEROS and NONPHOTOTROPIC HYPOCOTYL 4. *Development* 131:1089–1100
- Harrigan GG, Skogerson K, MacIsaac S, Bickel A, Perez T, Li X (2015) Application of 1H NMR profiling to assess seed metabolomic diversity. A case study on a soybean era population. *J Agric Food Chem* 63:4690–4697
- He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5:522–531
- Herman EM, Larkins BA (1999) Protein storage bodies and vacuoles. *Plant Cell* 11:601–613
- Hill JF, Breidenbach RW (1974) Proteins of soybean seeds II. Accumulation of the major protein components during seed development and maturation. *Plant Physiol* 53:747–751
- Höglund A-S, Rödin J, Larsson E, Rask L (1992) Distribution of napin and cruciferin in developing rape seed embryos. *Plant Physiol* 98:509–515
- Huang DQ, Koh C, Feurtado JA, Tsang EWT, Cutler AJ (2013) MicroRNAs and their putative targets in *Brassica napus* seed maturation. *BMC Genom*. doi:10.1186/1471-2164-14-140
- Huber SC, Hardin SC (2004) Numerous posttranslational modifications provide opportunities for the intricate regulation of metabolic enzymes at multiple levels. *Curr Opin Plant Biol* 7:318–322
- Hwang EY, Song Q, Jia G, Specht JE, Hyten DL, Costa J, Cregan PB (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genom*. doi:10.1186/1471-2164-15-1
- Islam F, Rengifo J, Redden R, Basford K, Beebe SE (2003) Association between seed coat polyphenolics (tannins) and disease resistance in common bean. *Plant Foods Hum Nutr* 58:285–297
- Jin D, Wang Y, Zhao Y, Chen M (2013) MicroRNAs and their cross-talks in plant development. *J Genet Genomics* 40:161–170
- Jofuku KD, Den Boer B, Van Montagu M, Okamoto JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6:1211–1225
- Jones SI, Vodkin LO (2013) Using RNA-Seq to profile soybean seed development from fertilization to maturity. *PLoS One*. doi:10.1371/journal.pone.0059270
- Joshi T, Yao Q, Franklin LD et al (2010) SoyMetDB: the soybean metabolome database. In: *Bioinformatics and biomedicine (BIBM)*, 2010. IEEE, pp 203–208
- Joshi T, Patil K, Fitzpatrick MR et al (2012) Soybean Knowledge Base (SoyKB): a web resource for soybean translational genomics. *BMC Genom*. doi:10.1186/1471-2164-13-S1-S15
- Kachroo P, Shanklin J, Shah J, Whittle EJ, Klessig DF (2001) A fatty acid desaturase modulates the activation of defense signaling pathways in plants. *Proc Natl Acad Sci USA* 98:9448–9453
- Katavic V, Agrawal GK, Hajduch M, Harris SL, Thelen JJ (2006) Protein and lipid composition analysis of oil bodies from two *Brassica napus* cultivars. *Proteomics* 6:4586–4598
- Kersten B, Agrawal GK, Iwahashi H, Rakwal R (2006) Plant phosphoproteomics: a long road ahead. *Proteomics* 6:5517–5528

- Kohler C, Lafon-Placette C (2015) Evolution and function of epigenetic processes in the endosperm. *Front Plant Sci*. doi:10.3389/fpls.2015.00130
- Korber N, Bus A, Li J, Parkin IA, Wittkop B, Snowdon RJ, Stich B (2016) Agronomic and seed quality traits dissected by genome-wide association mapping in *Brassica napus*. *Front Plant Sci*. doi:10.3389/fpls.2016.00386
- Korbes AP, Machado RD, Guzman F et al (2012) Identifying conserved and novel microRNAs in developing seeds of *Brassica napus* using deep sequencing. *PLoS One*. doi:10.1371/journal.pone.0050663
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*. doi:10.1186/1746-4811-9-29
- Kortseniemi M, Vuorinen AL, Sinkkonen J, Yang B, Rajala A, Kallio H (2015) NMR metabolomics of ripened and developing oilseed rape (*Brassica napus*) and turnip rape (*Brassica rapa*). *Food Chem* 172:63–70
- Lafon-Placette C, Köhler C (2014) Embryo and endosperm, partners in seed development. *Curr Opin Plant Biol* 17:64–69
- Lam HM, Xu X, Liu X et al (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* 42:1053–1059
- Le BH, Wagmaster JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB (2007) Using genomics to study legume seed development. *Plant Physiol* 144:562–574
- Lei MG, Reeck GR (1987) Two-dimensional electrophoretic analysis of the proteins of isolated soybean protein bodies and of the glycosylation of soybean proteins. *J Agric Food Chem* 35:296–300
- Li L, Hur M, Lee JY et al (2015) A systems biology approach toward understanding seed composition in soybean. *BMC Genom*. doi:10.1186/1471-2164-16-S3-S9
- Li D, Liu Z, Gao L et al (2016) Genome-wide identification and characterization of microRNAs in developing grains of *Zea mays* L. *PLoS One*. doi:10.1371/journal.pone.0153168
- Libault M, Farmer A, Joshi T et al (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J* 63:86–99
- Lin L, Allemekinders H, Dansby A, Campbell L, Durance-Tod S, Berger A, Jones PJ (2013) Evidence of health benefits of canola oil. *Nutr Rev* 71:370–385
- Lin H, Rao J, Shi J et al (2014) Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. *J Integr Plant Bio* 56:826–836
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133:523–536
- Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, Carrington JC (2007) Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. *Plant J* 52:133–146
- Liu X, Huang J, Wang Y, Khanna K, Xie Z, Owen HA, Zhao D (2010) The role of floral organs in carpels, an *Arabidopsis* loss-of-function mutation in MicroRNA160a, in organogenesis and the mechanism regulating its expression. *Plant J* 62:416–428
- Liu S, Liu Y, Yang X et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun*. doi:10.1038/ncomms4930
- Lorenz C, Rolletschek H, Sunderhaus S, Braun H-P (2014) *Brassica napus* seed endosperm—metabolism and signaling in a dead end tissue. *J Proteomics* 108:382–426
- M-a Ohto, Fischer RL, Goldberg RB, Nakamura K, Harada JJ (2005) Control of seed mass by APETALA2. *Proc Natl Acad Sci USA* 102:3123–3128
- M-a Ohto, Floyd SK, Fischer RL, Goldberg RB, Harada JJ (2009) Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*. *Sex Plant Reprod* 22:277–289
- Mallory AC, Dugas DV, Bartel DP, Bartel B (2004) MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Curr Biol* 14:1035–1046
- Mallory AC, Bartel DP, Bartel B (2005) MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* 17:1360–1375
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18:1509–1517
- Meinke D, Chen J, Beachy R (1981) Expression of storage-protein genes during soybean seed development. *Planta* 153:130–139
- Meyer LJ, Gao J, Xu D, Thelen JJ (2012) Phosphoproteomic analysis of seed maturation in *Arabidopsis*, rapeseed, and soybean. *Plant Physiol* 159:517–528
- Michael TP, VanBuren R (2015) Progress, challenges and the future of crop genomes. *Curr Opin Plant Biol* 24:71–81
- Miller SS, Bowman L-AA, Gijzen M, Miki BL (1999) Early development of the seed coat of soybean (*Glycine max*). *Ann Bot* 84:297–304
- Moïse JA, Han S, Gudynaite-Savitch L, Johnson DA, Miki BLA (2005) Seed coats: structure, development, composition, and biotechnology. *In Vitro Cell Dev Biol-PI* 41:620–644
- Morrell PL, Buckler ES, Ross-Ibarra J (2011) Crop genomics: advances and applications. *Nat Rev Genet* 13:85–96
- Mukherji M (2005) Phosphoproteomics in analyzing signaling pathways. *Expert Rev Proteomics* 2:117–128
- Nakabayashi R, Saito K (2013) Metabolomics for unknown plant metabolites. *Anal Bioanal Chem* 405:5005–5011
- Narvel JM, Fehr WR, Welke GA (1998) Agronomic and seed traits of soybean lines lacking seed lipoxygenases. *Crop Sci* 38:926–928
- Nielsen NC, Dickinson CD, Cho T-J et al (1989) Characterization of the glycinin gene family in soybean. *Plant Cell* 1:313–328
- Niu Y, Wu GZ, Ye R et al (2009) Global analysis of gene expression profiles in *Brassica napus* developing seeds reveals a conserved lipid metabolism regulation with *Arabidopsis thaliana*. *Mol Plant* 2:1107–1122
- Nodine MD, Bartel DP (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Gene Dev* 24:2678–2692
- Norton G, Harris JF (1975) Compositional changes in developing rape seed (*Brassica napus* L.). *Planta* 123:163–174
- Ohlrogge JB, Kuo T-M (1984) Control of lipid synthesis during soybean seed development: enzymic and immunochemical assay of acyl carrier protein. *Plant Physiol* 74:622–625
- Oikawa A, Nakamura Y, Ogura T et al (2006) Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. *Plant Physiol* 142:398–413
- Oikawa A, Matsuda F, Kusano M, Okazaki Y, Saito K (2008) Rice metabolomics. *Rice* 1:63–71
- Okushima Y, Overvoorde PJ, Arima K et al (2005) Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* 17:444–463
- O'Rourke JA, Bolon YT, Bucciarelli B, Vance CP (2014) Legume genomics: understanding biology through DNA and RNA sequencing. *Ann Bot* 113:1107–1120
- Pawson T, Scott JD (2005) Protein phosphorylation in signaling—50 years and counting. *Trends Biochem Sci* 30:286–290

- Penfield S, Meissner RC, Shoue DA, Carpita NC, Bevan MW (2001) MYB61 is required for mucilage deposition and extrusion in the Arabidopsis seed coat. *Plant Cell* 13:2777–2791
- Peng FY, Weselake RJ (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. *BMC Genom*. doi:10.1186/1471-2164-12-286
- Petolino JF (2015) Genome editing in plants via designed zinc finger nucleases. *In Vitro Cell Dev Pl* 51:1–8
- Porta H, Rocha-Sosa M (2002) Plant lipoxygenases. Physiological and molecular features. *Plant Physiol* 130:15–21
- Radchuk V, Borisjuk L (2014) Physical, metabolic and developmental functions of the seed coat. *Front Plant Sci*. doi:10.3389/fpls.2014.00510
- Ripp KG, Viitanen PV, Hitz WD, Franceschi VR (1988) Identification of membrane protein associated with sucrose transport into cells of developing soybean cotyledons. *Plant Physiol* 88:1435–1445
- Rubel A, Rinne R, Canvin D (1972) Protein, oil, and fatty acid in developing soybean seeds. *Crop Sci* 12:739–741
- Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Schultz DJ, Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis. In: *Lipid biotechnology*. CRC Press, pp 1–25
- Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y (2004) Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432:779–782
- Severin AJ, Woody JL, Bolon YT et al (2010) RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol*. doi:10.1186/1471-2229-10-160
- Shamimuzzaman M, Vodkin L (2012) Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. *BMC Genom*. doi:10.1186/1471-2164-13-310
- Shanklin J, Cahoon EB (1998) Desaturation and related modifications of fatty acids 1. *Annu Rev Plant Biol* 49:611–641
- Shewry PR (2009) Wheat. *J Exp Bot* 60:1537–1553
- Smith CW, Patton JG, Nadal-Ginard B (1989) Alternative splicing in the control of gene expression. *Annu Rev Genet* 23:527–577
- Song Q-X, Liu Y-F, Hu X-Y, Zhang W-K, Ma B, Chen S-Y, Zhang J-S (2011) Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol*. doi:10.1186/1471-2229-11-5
- Sreenivasulu N, Wobus U (2013) Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annu Rev Plant Biol* 64:189–217
- Swatek KN, Graham K, Agrawal GK, Thelen JJ (2011) The 14-3-3 isoforms chi and epsilon differentially bind client proteins from developing Arabidopsis seed. *J Proteome Res* 10:4076–4087
- Tan H, Xie Q, Xiang X et al (2015) Dynamic metabolic profiles and tissue-specific source effects on the metabolome of developing seeds of Brassica napus. *PLoS One*. doi:10.1371/journal.pone.0124794
- Thelen JJ, Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. *Metab Eng* 4:12–21
- Thompson R, Burstin J, Gallardo K (2009) Post-genomics studies of developmental processes in legume seeds. *Plant Physiol* 151:1023–1029
- Thomson DW, Bracken CP, Goodall GJ (2011) Experimental strategies for microRNA target identification. *Nucleic Acids Res* 39:6845–6853
- Trapnell C, Williams BA, Pertea G et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515
- Voelker T, Kinney AJ (2001) Variations in the biosynthesis of seed-storage lipids. *Annu Rev Plant Biol* 52:335–361
- Wang J-W, Wang L-J, Mao Y-B, Cai W-J, Xue H-W, Chen X-Y (2005) Control of root cap formation by microRNA-targeted auxin response factors in Arabidopsis. *Plant Cell* 17:2204–2216
- Wang X, Wang H, Wang J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43:1035–1039
- Weber H, Borisjuk L, Wobus U (2005) Molecular physiology of legume seed development. *Annu Rev Plant Biol* 56:253–279
- Weijers D, Schlereth A, Ehrismann JS, Schwank G, Kientz M, Jürgens G (2006) Auxin triggers transient local signaling for cell specification in Arabidopsis embryogenesis. *Dev Cell* 10:265–270
- Wu G, Zhang L, Yin Y, Wu J, Yu L, Zhou Y, Li M (2015) Sequencing, de novo assembly and comparative analysis of Raphanus sativus transcriptome. *Front Plant Sci*. doi:10.3389/fpls.2015.00198
- Xia Z, Xu H, Zhai J, Li D, Luo H, He C, Huang X (2011) RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. *Plant Mol Biol* 77:299–308
- Xu HM, Kong XD, Chen F, Huang JX, Lou XY, Zhao JY (2015) Transcriptome analysis of *Brassica napus* pod using RNA-Seq and identification of lipid-related candidate genes. *BMC Genom*. doi:10.1186/s12864-015-2062-7
- Ye CY, Xu H, Shen E et al (2014) Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean. *Front Plant Sci*. doi:10.3389/fpls.2014.00743
- Yu J, Zhang Z, Wei J, Ling Y, Xu W, Su Z (2014) SFGD: a comprehensive platform for mining functional information from soybean transcriptome data and its use in identifying acyl-lipid metabolism pathways. *BMC Genom*. doi:10.1186/1471-2164-15-271
- Zabala G, Campos E, Varala KK et al (2012) Divergent patterns of endogenous small RNA populations from seed and vegetative tissues of Glycine max. *BMC Plant Biol*. doi:10.1186/1471-2229-12-177
- Zhang Y, Liang W, Shi J, Xu J, Zhang D (2013) MYB56 encoding a R2R3 MYB transcription factor regulates seed size in *Arabidopsis thaliana*. *J Integr Plant Bio* 55:1166–1178
- Zhang H, Zhang J, Wei P et al (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol* 12:797–807
- Zhang XQ, Sun J, Cao XF, Song XW (2015) Epigenetic mutation of RAV6 affects leaf angle and seed size in rice. *Plant Physiol* 169:2118–2128
- Zhou Z, Jiang Y, Wang Z et al (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* 33:408–414