Systems biology of seeds: decoding the secret of biochemical seed factories for nutritional security

Anil Kumar^{1,2} · Rajesh Kumar Pathak^{2,3} · Aranyadip Gayen² · Supriya Gupta² · Manoj Singh² · Charu Lata⁴ · Himanshu Sharma⁵ · Joy Kumar Roy⁵ · Sanjay Mohan Gupta⁶

Received: 12 April 2018 / Accepted: 16 October 2018 / Published online: 24 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Seeds serve as biochemical factories of nutrition, processing, bio-energy and storage related important bio-molecules and act as a delivery system to transmit the genetic information to the next generation. The research pertaining towards delineating the complex system of regulation of genes and pathways related to seed biology and nutrient partitioning is still under infancy. To understand these, it is important to know the genes and pathway(s) involved in the homeostasis of bio-molecules. In recent past with the advent and advancement of modern tools of genomics and genetic engineering, multi-layered 'omics' approaches and high-throughput platforms are being used to discern the genes and proteins involved in various metabolic, and signaling pathways and their regulations for understanding the molecular genetics of biosynthesis and homeostasis of bio-molecules. This can be possible by exploring systems biology approaches via the integration of omics data for understanding the intricacy of seed development and nutrient partitioning. These information can be exploited for the improvement of biologically important chemicals for large-scale production of nutrients and nutraceuticals through pathway engineering and biotechnology. This review article thus describes different omics tools and other branches that are merged to build the most attractive area of research towards establishing the seeds as biochemical factories for human health and nutrition.

Keywords Bio-fortification · Nutraceuticals · Nutrients partitioning in seeds · Seed biology · Systems biology

Introduction

Nutritional systems science deals with food and nutrient requirements for heterogeneous populations. Among different nutritional components, seeds are one of the key materials of nutraceutical and pharmaceuticals resources consisting of all possible kinds of nutritionally important and bioactive molecules that include various soluble carbohydrates, storage proteins, starch polymer and lipids needed for our daily diet. Apart from this, they also function as an important delivery system of genetic information from generations to generation. These stored seed reserves comprise 70% of total caloric intake in the form of food and animal feed produced through sustainable agriculture (Sreenivasulu 2017). Therefore, seeds of appropriate characteristics are required to meet the demand of diverse agro-climatic conditions and intensive cropping systems.

Further, the nutrient partitioning phenomena in the seed tissues is a new area of interest to improve the quality of seeds as it observes specific differentiation of seed tissues into embryo, endosperm, aleurone layer and accordingly,

Anil Kumar, Rajesh Kumar Pathak and Aranyadip Gayen contributed equally.

✉ Anil Kumar
anilkumar.mbge@gmail.com

- ¹ Rani Lakshmi Bai Central Agricultural University, Jhansi, Uttar Pradesh 284003, India
- ² Department of Molecular Biology and Genetic Engineering, College of Basic Sciences and Humanities, G. B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand 263145, India
- ³ Department of Biotechnology, G. B. Pant Institute of Engineering and Technology, Pauri Garhwal, Uttarakhand 246194, India
- ⁴ Council of Scientific and Industrial Research-National Botanical Research Institute, Lucknow, India
- ⁵ National Agri-Food Biotechnology Institute, Mohali, Punjab 140306, India
- ⁶ Molecular Biology and Genetic Engineering Laboratory, Defence Institute of Bio-Energy Research (DIBER), DRDO, Haldwani 263139, India

partitioning of the different nutritionally important bio-molecules (Nadeau et al. 1995). The previous studies in the area of conventional genetics and breeding have large contribution towards the identification of factors involved in crop growth and grain filling. Apart from genetic and epigenetic factors, environment also plays important role in seed development. The major bio-molecules like proteins, carbohydrates, lipids, etc. are distributed across the seed tissues, however, it is crucial to discern if nutrient partitioning is influenced by these factors. Nevertheless, the genetic factors should be explored at molecular level for understanding seed biology (Xie et al. 2015; Jing et al. 2016).

Different 'omics' based platforms are the best way to understand seed biology at molecular to systems level as it would provide holistic views of the complex biological phenomenon of the seed-system by interpreting the behavior of each constituent (gene, protein and metabolites) and their interactions in seed. It can also identify various novel properties that can be further utilized in seed biology research programme (Fukushima and Kusano 2013; Kumar et al. 2015a, b). This review thus provides molecular insights for understanding the intricacy of seed biology and nutrient partitioning.

Factors influencing seed developmental machinery

Environment

Environmental factors such as low water or herbivory may change the maternal influences on seed development. They can decrease the resources available for the development of seeds as well as can also affect the key regulatory event during and after pollination (Gehring and Delph 2006; Diggle et al. 2010). The lamination of resources during seed development may cause maternal plants to make smaller seeds. While, the molecular mechanisms of maternal control are not characterized well, the paternal part also influences the developmental mechanisms of seeds in two ways; in first way, a few embryos might develop quicker because of correlations among paternal genes that affect mating success and growth of offspring; in second way, paternal alleles contribute to embryo as well as endosperm vigour that results in variation in mature embryo and seed size (Kigel 1995).

Nutrition

The translocation of nutrients during seed development into seed coat, embryo or endosperm is yet not well defined. Moreover, only a few members of plant species have been studied in details. The literature studies showed that unusual transport mechanisms are occupied by plants for the

nutrition of embryo as there is no vascular connection between the vascular bundle and embryo. The transport of nutrients may occur with the help of symplast and apoplast pathways. The O₂ and CO₂ may be transported with the help of intercellular spaces of the parenchyma middle layer (Kigel 1995). Pathway associated with seed biology and their role in nutrition, i.e. amino acid biosynthetic pathways, glycolysis, gluconeogenesis, pyruvate and energy metabolism are reported. Besides, hormone signaling pathway such as jasmonate biosynthesis (JA), IAA/ethylene and gibberellin are also associated with seeds (Sreenivasulu and Wobus 2013; Li et al. 2015). Advances are being made towards finding the key components of the seeds that controlling nutrient loading during their developmental process through systems biology (Zhang et al. 2007; Kumar et al. 2015a, b).

Physiology

Deeper understanding of the plant physiological parameters is key concepts in developing molecular biology background to visualize seed development. Micromanipulation and video microscopy have been used to observe single fusion pairs. Genes involved in pollen tube guidance or pollen discharge in synergids, as well as genes exhibiting differential expression in sperm, egg and central cells before and after fertilization have been identified (Raghavan 2000). The coordination of development, differentiation and maturation processes is based on the constant transmission and perception of signals by seed coat, endosperm and embryo. Also, the ratio of phytohormones in different signal regulates particular steps in seed development (Locascio et al. 2014). Further, several technologies, for example, infrared thermography have been developed that can decipher biophysical and biochemical changes linked with imbibition, germination and overall seed vitality (Kranner et al. 2010; Macovei et al. 2017). Interestingly, reactive oxygen species (ROS) are known to play important role during seed development both as signalling and damaging molecules (Kumar et al. 2015a, b). A strong interconnection between ROS, phytohormones and DNA repair has been envisaged at all seed developmental stages right from embryogenesis to germination (El-Maarouf-Bouteau et al. 2013; Kumar et al. 2015a, b). However, hydration state of the seed is also a major factor that condition ROS signalling and DNA repair (Bewley et al. 2012).

ROS as signalling molecules are involved in different signalling pathways during seed imbibition and germination, for example, mitogen-activated protein kinases and pentose phosphate pathway (El-Maarouf-Bouteau et al. 2013; Diaz-Vivancos et al. 2013). Furthermore, several antioxidative enzymes are reportedly inhibited or induced during seed imbibitions (Balestrazzi et al. 2011). On the other hand, several phytohormones are known to influence ROS production during seed development and germination as for example,

ABA resulted in decreased ROS accumulation in barley (Ishibashi et al. 2012), rice and sunflower (Zhang et al. 2013) while gibberellic acid led to increased ROS generation in *Arabidopsis* (Lariguet et al. 2013). Considering these crucial information, it is thus an essential requirement that omics techniques must be effectively exploited to decipher various complex mechanisms involved in seed development.

Epigenetic modifications

Genomic imprinting is an autonomous and independent phenomenon that occurs in flowering plants and mammals. Placenta and endosperm imprinting occurs at the site of embryo-nourishing tissues, and imprinted genes are found to be responsible for the regulation of translocation of nutrients to the developing progeny. The regulation of many imprinted plant genes is highly influenced by the antagonistic actions of DNA methylation and Polycomb group-mediated histone methylation (Jiang and Köhler 2012). The genes with imprinted expression are conserved between dicots and monocots suggesting that long-term selection can maintain imprinted expression at some loci (Ohnishi et al. 2014). In *Arabidopsis thaliana*, genome dosage is controlled by a coordination of the genetic pathway, HAIKU (IKU), with epigenetic controls. Actually, DNA methylation and trimethylated lysine 27 on histone H3 (H3K27me3) deposition determine the seed size (Li et al. 2013). Recently regulatory roles of DNA methylation in rice seed maternal integument development have been elaborated wherein whole genome bisulfite deep sequencing approach was used to identify differentially methylated regions and genes underlying this crucial phenomenon (Wang et al. 2017). In another interesting study, it was revealed that a seed-specific transcription factor gene LEAFY COTYLEDON 1 reverses the silenced state inherited from gametes by promoting initial establishment of an active chromatin state and activating its expression de novo at FLOWERING LOCUS C (Tao et al. 2017).

Small regulatory RNA

Small non-coding RNAs including microRNAs (miRNAs) and the short interfering RNAs (siRNAs) are key monitors of gene expression at transcriptional and post-transcriptional levels and are also reported to play vital roles in regulation of seed developmental processes and germination (Rodrigues and Miguel 2017). The biological impact of the presence of different siRNA profiles in seeds is not yet clear since studies have been focused mostly on miRNAs which are generally 21-nt long (Rodrigues and Miguel 2017).

The down-regulation of target genes by miRNA plays a critical role in seed development and germination (Rogers and Chen 2013). Differential expression of miRNAs in tissues has been demonstrated in *Arabidopsis* mutant

(Zhai et al. 2014). Artificial mi-RNA (ami-RNA) has been designed and expressed targeting specific gene and thereby ensured defective embryogenesis. Dicer Like-1 (DCL-1) gene is responsible to catalyze the formation of proper structure from secondary precursor (Nodine and Bartel 2010). Similarly, *emb76* a DCL-1 defective mutant, exhibited embryo arrest and failed to suppress differentiation of a special type of cell known as ‘suspensor’ (Bewley et al. 2012). Further analysis of *dcl1-5* and *dcl1-10* mutants suggested that DCL1 mutations have an early impact on 8-cell embryonic stage, most probably due to the up-regulation of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) 10 and SPL 11 which are targeted by miR156 (Xu et al. 2016).

From the point of view of miRNA biogenesis and maturation, the involvement of methyl transferase has been reported (Pluskota et al. 2011). miRNAs are also crucial regulators of the timing of seed maturation (Willmann et al. 2011). They usually repress the positive regulators of seed maturation program during early embryogenesis which was confirmed by the upregulation of *FUS3*, *LEC2*, *L1L*, *NF-YB6*, *MYBs* and *bZIPs* and down-regulation of *ARABIDOPSIS 6B-INTERACTING PROTEIN1-LIKE1* (*ASIL1*) and *ASIL2* and histone deacetylase *HDA6/SIL1* in *dcl15* mutants. Further, miRNAs have also been implicated in seed dormancy and transition to germination. miR156 and miR172 have been implicated in affecting seed germination wherein miR156 inhibited germination in lettuce and *Arabidopsis* while miR172 promoted the same (Huo et al. 2016). Additionally, miR159 which targets *MYB33* and *MYB101* is known to positively regulate ABA responses important for the transition from seed dormancy to germination in *Arabidopsis* (Nonogaki 2017). Besides, recent study conducted on rice plant demonstrated the presence of alternative splicing events of coding and long non-coding RNA (lncRNA) during the developmental stage of seed via comparison of immature seed with embryo as well as endosperm of mature seed. It was predicted that the developing seeds had additional alternative splicing event, i.e. 5.8–57 time more in comparison to root, leaves, buds, flowers and reproductive meristems. Therefore, the alternative splicing of lncRNA need further study to decode its intricate processing during seed development (Kiegle et al. 2018).

Molecular status for decoding the complexity of seed biology and nutrients partitioning

The studies on seed development have emerged as an important area of research in plant developmental biology (Lohe and Chaudhury 2002). The processes of seed development are regulated by a number of genes and

their interactions with its target. The main regulators of seed development contain B3 or HAP3 domains, which interact with basic leucine zipper (bZIP) and APETALA2 (AP2) transcription factors (TFs). Other TFs that play an indispensable role during this process contain homeobox, NAM, ATAF and CUC (NAC), Myeloblastosis (MYB), and Auxin Response Factor (ARF)-domains (Agarwal et al. 2011). Many genes viz., MYB (Dumas and Rogowsky 2008), LEAFY COTYLEDON2 (LEC2) (Stone et al. 2001), GNOM (Grebe et al. 2000), SHOOT MERISTEMLESS (STM) (Long and Barton 2000), MONOPTEROS (Hardtke and Berleth 1998) and FACKEL (Schrack et al. 2000) are playing critical role, and also LEC2 is likely to be a transcriptional regulator which provides the correct cellular environment for the initiation of embryo development. ALTERED MERISTEM PROGRAMMING1 (AMP1) gene was identified and has vital role in embryogenesis during pattern formation (Lohe and Chaudhury 2002) and LEAFY COTYLEDON (LEC) genes *LEC1*, *LEC1-LIKE*, *LEC2*, and *FUS3* are known to

be seed-specific TF genes involved in the regulation of seed developmental processes (Le et al. 2010) (Fig. 1).

The developmental mechanisms of seeds are different in monocotyledonous and dicotyledonous plants, the analysis of high-throughput transcriptome data provide first insight in deciphering the molecular networks and pathways that function during development in individual compartments of seeds. Comparative systems based coexpression network analyses define the evolutionarily conservative (*FUS3/ABI3/LEC1*) as well as divergent (*LEC2*) networks in dicots and monocots (Sreenivasulu and Wobus 2013).

Homeotic genes have key roles in differentiation and development of different plant parts. For example, the regulation of floral homeotic gene expression is controlled by AP2 in *Arabidopsis*. In consistency with its genetic functions, it has been reported that AP2 is expressed at the RNA level in all four types of floral organs including sepals, petals, stamens, and carpels and in developing ovules (Jofuku et al. 1994). In particular angiosperm species it has been shown that B-sister genes are important for correct

Fig. 1 Gene controlling nutrient distribution and partitioning during seeds development

differentiation of the ovule/seed. Still, numbers of genes particularly expressed in the seed are under characterization or uncharacterized stage and have unknown functions. Hence, there are more challenges for characterization of seed-specific genes that may be used in biofortification program to securing food and nutritional security for our society. Therefore, modern systems biology approaches can have significant benefits toward complete knowledge of seed biology.

Potential promises of omics approaches

Potential of ‘omics’ approaches open the gates for novel technologies, which would help in taking the advantage by studying different components of biological systems that are required for proper functioning of the cell. It is expected that the applied integrative omics-based approaches may be helpful in understanding the biology of seeds and its developmental processes (Fig. 2). These ‘omics’ based approaches include the incorporation of genomics, transcriptomics, proteomics, metabolomics, glycomics, lipidomics and other omics sciences to perform the studies on biological

molecules involved in the appropriate working of the cells and its physiological mechanism and their responses with respect to different environmental changes (Kumar et al. 2015a, b).

Phenomics: characteristics features of seeds

In future, phenomics having simulation approach, which help to get unbiased data of biological systems is going to play an important role in understanding the intricacies of seed biology (Navarro et al. 2016). Seeds have complex phenotypes which can be difficult to assess quantitatively (Gustin and Settles 2015). Novel approaches are essential now a day to harness quantitative phenotypes and to explain the genetic basis of agriculturally important traits. The screening of germplasm with high-performance characteristics would be thus easier in resource-limited environments. Recently, plant phenomics has offered and integrated new technologies to improve the description of complex plant phenotypes. With the growing high-throughput phenotyping platforms it has been also possible to capture phenotypic data from plants in a non-destructive manner (Rahaman et al. 2015).

Fig. 2 Omics approaches for deciphering the secret of biochemical factories in seeds

Recently, researchers have applied computational algorithms for deciphering the complexity of the phenomics data such as progression of plant senescence by the analysis of color images, both distorted and blurred under high throughput conditions (Cai et al. 2016) as well as phenomic prediction of maize hybrids to get viable alternative for genomic and metabolic prediction of the hybrid performance (Edlich-Muth et al. 2016).

Genomics and transcriptomics

The availability of the reference genome sequence of the model plant *Arabidopsis* and other crops of economic and agronomic importance has made it possible to develop tools that aid system-level integration of genes into functional units in large interconnected biological networks and accelerate crop improvement (Venglat et al. 2014; Gupta et al. 2017). The comparison of transcriptional networks across species that operate during embryo and seed development is now available by system biology approaches, thus providing deeper insights into conserved and divergent gene networks (Bevan and Uauy 2013). The detection of various seed coat genes, the developmental pathways controlling seed coat development, etc. is not completely explored and a global genetic program associated with seed coat development is still not reported. Global expression datasets that could be used for the identification of new candidate genes for seed coat development were generated. These dataset provided a comprehensive expression profile for developing seed coats in *Arabidopsis*. It should provide a useful resource and reference for other seed systems (Dean et al. 2011). Very recently transcriptomic analysis of seed coats in *Brassica napus* revealed novel potential TFs involved in proanthocyanidin biosynthesis thus providing pivotal information for the identification of candidate genes for seed coat color (Hong et al. 2017).

The global genetic and metabolic programs involved in seed development and contribution in genetic improvement of the seed is yet to be fully exploited in crop plants. By developing insights into molecular and biochemical programs associated with gene expression, protein and metabolite profiles during seed development in model and crop plants, the study is being proceeded to advanced level. These integrated systems approaches and studies are producing comprehensive datasets (Venglat et al. 2014). It has been already reported that seed size is regulated by genetic factors that have been under constant selection during the evolution of diverse seed systems in flowering plants (Westoby et al. 2002). Recently transcriptome of 35 wheat genotypes analyzed using RNASeq at two seed developmental stages namely, 14 days post anthesis (DPA) and 30 DPA revealed that flour yield in milling and bread quality were largely influenced due to expression patterns of genes associated

with flour yield at the earlier seed development stage (Rangan et al. 2017). Further, to explore the role of transcription factors in the transcriptional regulation of grain filling, the expression analysis of several known and putative transcription factor genes was done during grain filling. Cluster analysis discovered a group of transcription factor genes having expression profiles similar to that of reported grain filling genes. This group includes genes namely *RITA*, a bZIP type transcription factor gene those which highly expressed in aleurone and endosperm tissues and may be important in regulating gene expression in developing rice grains. Other genes included in this cluster are *Daf* genes, whose gene products can activate the expression of seed storage protein genes by binding to the prolamin-box, and nine known and putative *MYB* genes (Kumar et al. 2009; Gaur et al. 2011; Gupta et al. 2011, 2017, 2018).

By transcriptomic analysis, it has been revealed that almost 5% of the differentially expressed genes influence hormone response and metabolism during fruit set. Aux-IAA and auxin transcription factor (ARF) genes showed higher variations regarding expression levels during fruit development, i.e. these transcriptional regulators play an active role in the developmental process (Wang et al. 2009). Singular value decomposition (SVD) analysis of microarray data proved as an alternate method for measuring gene expression pattern recognition. SVD is very much useful in finding key players of grain filling and measured by comparison with results obtained via clustering 84% of these known grain-filling genes were re-identified by doing SVD analysis (Anderson et al. 2003). Changes in fatty acids and protein levels in developing soybean embryo have been also estimated. Fatty acids from hydrolyzed lipids and proteins were analyzed by GC-FID, and fluorescent hydrophobic protein assay, respectively (Collakova et al. 2013).

The endosperm and seed coat growth have also a significant influence on determining the final seed size. AP2 and MADS-box transcription factors play important functions regulating the seed size of *Arabidopsis*. A delayed cellularization of the endosperm in *ap2* mutant results in larger embryo sacs and bigger embryos that show increased cell number and size and this larger seed trait was passed through the maternal sporophyte and endosperm genome (Fang et al. 2012). Polycomb (PcG) protein-encoding genes are the second group of genes reported to have a role in controlling seed size. This includes genes that form the FERTILIZATION INDEPENDENT SEED (FIS) complex which functions as transcriptional repressors via histone methylation and imprinting genes (DNA METHYLTRANSFERASE1 and DECREASE IN DNA METHYLATION2) that methylate the PcG genes (Hennig and Derkacheva 2009). The third category of genes that regulate seed size includes *HAIKU* (*IKU*) and *MINISEED3* (*MINI3*) that are activated by *SHORT HYPOCOTYL UNDER BLUE1* (*SHB1*). Although

very little is known about the molecular processes that regulate endosperm cellularization in the monocot crops such as maize, rice and wheat, their increased nuclear proliferation of the early endosperm correlates with increased seed size, sink strength and grain weight (Figueiredo and Köhler 2014).

A specific expression of *KLUH*, a cytochrome P450 encoding gene, influences these events much in inner integument of developing ovules. *TT16* and *AGL63*, the two B-SIS-TER MADS box genes that have roles in female reproductive organ development in *Arabidopsis*, have been shown to be negative regulators of seed coat differentiation. Mutations in them result in larger seed size (Prasad et al. 2010). *Arabidopsis* encodes a ubiquitin receptor, *DA1*, which regulates the final seed and organ size thus impacting the period of cell proliferation. Mutation in *DA1* reduces seed size, whereas its over-expression leads to larger seeds. Genetic analysis has provided valuable insights into the transcriptional regulation of seed storage metabolism, for example, the role of *WRINKLED 1* (*WRI1*), *WRI3* and *WRI4* in the regulation of fatty acid synthesis in several plant species (Li et al. 2008).

Loss and gain of gene function

Technological advancement in functional genomics tools viz loss and gain of function in mutant provides a tremendous platform to explore the biology of seeds because all the genetic information of an organism are coded in its genome, and when the mutation occurs in the genome by changing its genomic sequences, structural and functional changes occur in proteins leading to diverse functions. Now a day's Clustered Regularly Interspaced Short Palindromic Repeats-Cas9 (CRISPR-Cas9) technology has revolutionized the investigations for the functions of essential genes very efficiently (Rajendran et al. 2015) and also advances in molecular systems biology and several other 'omics' platform play crucial role in identification and validation of those genes, which losses their function or confer a new function to understand molecular machinery of any systems or system of systems such as seed system or whole crop plants systems. Very recently this technology was used to knock out *EPLF9* which is a positive regulator of stomatal development (Hunt et al. 2010). Thus, the future scope of seed biology would be to characterize functional mutants of rate-limiting enzymes and key regulatory targets of seed storage components, as well as to revalidate the models for engineering seed sink for translational nutritional and health benefits (Sreenivasulu 2017).

RNAi and miRNA technologies used in gene silencing are newly developed tools. These have great advantages over antisense and co-suppression as they have higher silencing efficiency and shortened time period for screening the targeted plants. These technologies are very much useful for

gene or pathway discoveries through nutritional genomics, transcriptomics, proteomics and metabolomics in plants to improve human health (Tang et al. 2007). To explore the developmental program of seeds, different experiments are performed to investigate the metabolic pathway and corresponding anabolic biosynthetic genes. Major technologies that are useful in pathway discoveries namely, SIDEMAP (Stable Isotope-base based Dynamic Metabolic Profiling), MIA (Mass Isotopomer Analysis), microarray and gene silencing are the emerging ones.

Quantitative trait loci (QTLs) and genome wide association studies (GWAS)

In the post genomic era, quantitative trait loci analysis (QTLs) has been a precise and powerful approach to determine loci and genes that manage important and complex traits associated with seed development (Macovei et al. 2017). Several QTLs linked to seed dormancy have been investigated in wild and cultivated cereals such as rice (Wan et al. 2005; Gu et al. 2008), barley (Hori et al. 2007; Sato et al. 2016) and wheat (Imtiaz et al. 2008; Ogonnaya et al. 2008). The identified 183 seed oil QTLs in soybean, which was annotated as part of Soybase database (Grant et al. 2009), providing key information related to seed development and associated molecular mechanism to set up soybean as a model legume crop (Gupta et al. 2017). In a very interesting study Sudarshan et al. (2017) characterized "D" locus *Linum usitatissimum* (flax) by QTL mapping and identified an underlying *FLAVONOID 3'5' HYDROXYLASE (F3'5'H)* gene which determine the seed and flower color. A single nucleotide deletion in this gene truncates the deduced product thus leading to yellow seed and white flower phenotypes in flax (Sudarshan et al. 2017).

While, genome-wide association studies (GWAS) recently became more popular, because it defeated several limitations of QTLs analysis, permitting a quantitative determination of the association between each and every genotyped markers with a phenotype of interest (Macovei et al. 2017). GWAS analysis conducted on soybean identified significant associations with concentrations of seed protein for 40 single nucleotide polymorphisms (SNPs), which are located on different regions across 10 chromosomes, besides, important associations with seed oil content was also investigated for 25 SNPs found on different regions across 12 chromosomes (Hwang et al. 2014). A recent study exploited a GWAS based approach on *Brassica napus*, identified 17 important associations for seed glucosinolate content along with five associations for hemicellulose content of seed using 89 genotypes with ~4000 SNP markers (Gajardo et al. 2015). In another study, 112 seed quality associated SNPs were also identified by utilizing genotypic and phenotypic data for 12 seed quality traits linked with oil content, linoleic acid concentration

and oleic acid concentration through GWAS based approach (Körber et al. 2016). Therefore, this one is another systems based approaches that having immense potential for decoding the intricacy of seed for improving its nutritional quality.

Proteomics

Dynamic resolution of seed protein samples is extremely limited due to the presence of high-abundance storage proteins. Several methods have been developed during the past decade are available for protein extraction (Gupta and Misra 2016) and interventions of proteomics approaches will be helpful in understanding the seed biology in greater details. The key proteomics tools like MALDI-TOF, Edman sequencing, Q-TOF, LC-MS/MS, etc. become the most crucial player in the analysis of proteins (Deng et al. 2013). Proteomics is particularly helpful for crops as it may give a glimpse not only for nutritional values, but also about yield and how different factors are influenced by adverse conditions (Salekdeh and Komatsu 2007). Recently, a detailed proteomic analysis of rice leaf, root and seed have been done by tissue using two independent technologies, i.e. 2DE followed by LC-MS/MS and MudPIT, lead to the discovery of large numbers of novel proteins (Ramalingam et al. 2015). Many of these proteins were detected in all of the tissues and have a role in the central metabolic pathways, transcription control and mRNA biosynthesis and protein biosynthesis. The majority of proteins showed a tissue-specific expression pattern and involvement in metabolic pathways. They were visualized on a metabolic map to illustrate the contribution of proteins to tissue-specific metabolic pathways (Koller et al. 2002). Recently proteome profile of seed and pericarp of healthy guarana fruits indicated the existence of several stress response and defense-related proteins which is highly unlikely for fast growing and differentiating reproductive tissues (Souza et al. 2017). The proteome changes in rice seeds during different day after fertilization interval has been reported (He and Yang 2013). The presence of salt-soluble, urea-soluble fraction of proteins has been observed during endosperm, embryo development as well as in stress development. Flour quality of cereals is highly correlated with protein composition and functional quality, thus proteomics approach is very good for identifying flour making protein markers for suitable cultivars. The protein analysis of wheat kernels for amphiphilic proteins has resulted in finding physiological functions of these proteins in wheat kernels (Miernyk and Hajduch 2011). Further, 2-DE approach has extended its role in the identification of soluble proteins that play a significant role in stabilizing the gas bubbles in dough and influencing the crumbling structure of proteins. Proteomics has also helped in the construction of proteome map investigating the level of protein modification during barley malting and detecting the proteins associated with

beer quality. Proteomic analysis of rice leaf, root, and seed showed the presence of many allergenic proteins in the seeds, which implicate the uses of proteomic analysis of foods for the presence of allergens (Eldakak et al. 2013).

Metabolomics

Metabolomics has grown greatly as an invaluable diagnostic tool for biochemical phenotyping of biological systems. In combination, transcriptomics and metabolomics analysis indicated similar changes in the expression levels of metabolites and transcripts during different environments. The metabolites related to nitrogen metabolism, such as asparagine, γ -aminobutyric acid and allantoin were decreased in both low temperature (15 °C) as well as low nitrate maturation environments (He et al. 2016). Correspondingly, *ALLANTOINASE*, *NITRITE REDUCTASE 1*, *NITRATE REDUCTASE 1* and *NITRILASE 4* are the nitrogen metabolism genes which were differentially regulated at the low temperature and nitrate maturation environments, as compared to control conditions (He et al. 2016). *DELAY OF GERMINATION (DOG) 1* is a seed expressed gene required for the induction of dormancy. Transcriptomics and metabolomics studies uncover the additional role of *DOG1* and hypothesize that the function of *DOG1* is not limited to dormancy; but also play multiple roles during seed maturation by interfering with the ABA signaling components in *Arabidopsis thaliana* (Dekkers et al. 2016). Recently, extensive genetic variation in the metabolite abundance was detected during tomato seed germination indicating that metabolic composition is related to germination phenotypes and overall seed performance (Kazmi et al. 2013).

Over the past decades, a large amount of information related to mass spectra, compound names and structures, statistical/mathematical models, and metabolic pathways and metabolite profile data have been developed and stored in the form of databases. Such databases are complementary to each other and support efficient growth in metabolomics, although the data resources remain scattered across the World Wide Web. Available metabolome databases help to summarize the present status of development of related tools with particular focus on the plant metabolome. Data sharing would find ways for the robust interpretation of metabolomics data and advances in plant systems biology. Such large amount of data shall be utilized to fetch the key information with respect to seed systems and which would be helpful in engineering of metabolic pathways responsible for production of the nutrients, besides, it can also be utilized to develop strategies to increase or decrease the synthesis of high-energy compounds or manufacturing of any other nutrients in the seed as per our need (Fukushima and Kusano 2013; Fukushima et al. 2014; Kumar et al. 2015a, b).

Lipidomics

Lipidomics aims to quantitatively represent the different classes of lipid along with their molecular species in the biological systems. Advances in lipidomics facilitated the quantitative lipid analysis with a unique level of sensitivity and precision. The power of MS based lipidomics play crucial role in addressing the complexity of queries related to cell biology (Brügger 2014; Horn and Chapman 2014). Recent studies on the model plant *Arabidopsis* transgenic seeds have given insights into where dihydroxy ascorbate (DHA) accumulated and united with other fatty acids of neutral and phospholipids from the developing as well as mature seeds (Zhou et al. 2014). The lipidomics approaches have the potential for dissecting the intricacy of seed-biology at molecular to systems level for quantification of seed developmental process that can be utilized for increasing nutritional content of seeds through genetic engineering.

Glycomics–thermodynamics approach

Glycomics is the systemic study of all the glycan structure present in given cell, tissue or organism. It is one of the emerging technologies in the post-genomic era having immense potential for characterization of biological systems. Studies conducted on *N*-glycan structures of lotus seed using mass spectrometry-based glycomics and glycoproteomics techniques identified 19 *N*-glycan structure which was having high mannose (~20%), paucimannosidic (~40%) as well as complex forms (~40%) and lotus convicilin storage protein 2 (LCP2), having high mannose *N*-glycans which serve as a model system for deciphering the role of seed proteins and its glycosylation in food allergy (Dam et al. 2013). The seeds of plants having rich content of biomolecule such as carbohydrates that have complex information of carrying biopolymers are becoming an area of increasing interest in the scientific community (Hu et al. 2015). The dramatic changes in glycosylation have been seen in various key events such as embryogenesis and differentiation and it may also play a major role during the development of seed. Therefore, understanding the role of glycan during seed development may be helpful for identification of key molecules that can be utilized for understanding the complex biology of seed and the developmental processes thereof.

Vitamin analysis

Vitamin analysis is also possible like the other molecules present in the seed tissues by several detectors as per its biochemical configurations. An ion-pair, reversed-phase HPLC separation method coupled with ESI-MS/MS detection system allows simultaneous determination of water-soluble vitamins. It is further proven that it has been used efficiently

for the extraction of total vitamin content from foodstuff. It is also proven that applying 0.05% (v/v) heptafluorobutyric acid as ion-pairing reagent in the mobile phase can ensure the separation of: thiamine (B_1), riboflavin (B_2), nicotinic acid, nicotinamide, pyridoxine (PN), pyridoxamine (PM), piridoxal (PL), thiamine-monophosphate (TMP), riboflavine-5'-phosphate (FMN) and the piridoxal-5'-phosphate (PLP) (Engel 2009). These methods identify the various vitamins and their concentration present in the seed; which can further be useful for the development of agri-food products with high nutritional values.

Minerals analysis

The presence of different micro and macro nutrients in the seed tissues are crucial source of nutrition to the human body. Analysis of minerals is a common practice in the agricultural research sector. The *in vivo* mineral distribution profiles in rice grains and shifts in those distribution patterns during advanced stages of germination were analyzed by synchrotron X-ray microfluorescence (XRF). Translocation of the minerals from specific seed locations during germination was also element specific. It was observed that high mobilization of K and Ca from grains to growing roots and leaf primordia and the flux of Zn to these expanding tissues were comparatively less than that of K and Ca. The mobilization of Mn or Fe was relatively low, at least during the first few days of germination (Lu et al. 2013).

Systems biology: a holistic approach for integration and analysis of seed biology data

A system includes a set of components or elements such as genes, proteins, metabolites, etc. in any cell. These components of a system are interconnected and interdependent in a structured way to facilitate the flow of information directly or indirectly, to maintain the balance of system, which are essential for existence and achieving its goal. This holistic approach to investigate its behavior through integration, annotation, modeling and analysis of high-throughput omics data generated from available omics platform is known as systems biology (Pathak et al. 2013; Kumar et al. 2015a, b).

Advances in high-throughput 'omics' technologies have generated a huge amount of biological data related to seed, and significant efforts have been made to understand biological systems as true systems for a cost-effective and affordable systems biology study. Analysis of high-throughput omics data through bioinformatics approaches involving database handling, modeling, network analysis and simulation results tremendous progress in system-level understanding of biological systems (Kumar et al. 2015b; Gupta and

Misra 2016; Pathak et al. 2017). Systems biology is essential to understand the molecular basis of seed from molecular to cellular and systems level for dissecting the intricacy of each component and its specific role in developmental processes and assist to increasing of the nutrient production in seeds through bio-fortification, metabolic engineering and other biotechnological approaches that will definitely help in nutritional security (Fig. 3).

The development of embryos in the seeds is a large scale metabolic conversion processes. This process imports photosynthates and amino acid precursors, which are converted into oil, protein and storage polysaccharides. The time period of such developmental processes is genetically pre-calculated and controlled by environmental factors (Li et al. 2015). The current biological networks models are based on accumulation data, which cannot predict the complete picture of cellular metabolism (Gurwitz 2014). Previous omics based studies through systems biology tools have investigated the set of specific genes and their molecular mechanisms through combined analysis of metabolomics and transcriptomics to decode the global developmental and

metabolic networks that determined the structural features and investigated the biochemical composition of the mature soybean seed (Li et al. 2015). Such types of studies may help in the identification of nutritional protein found in seeds for curtailing hidden hunger, malnutrition and maintaining health and development of nutraceuticals product for the benefits of society.

Integrative seed systems biology

It begins with sample collection and wet lab experimentation using hi-throughput omics platforms such as genomics, transcriptomics, phenomics, metabolomics, lipidomics, glycomics and statistical data analysis through execution of computational tools, data integration and interpretation as well as knowledge discovery. It tries to build the integrative models based on data analysis and organization of systems such as seed systems at molecular level with emphasizes precise queries at specific scale. This approach is also known as ‘Top-down’ (Kumar et al. 2015a, b; Gupta and Misra 2016).

Fig. 3 Systems biology approaches for deciphering the complexity of seed and its developmental process at molecular to systems level

Recent developments in omics platforms have started to provide the information of some key genes/metabolites that are required for the development of seed. Analyses and visualization of mRNA abundance during seed developmental processes, maturation, and beyond are providing evidences as to the global gene expression which determines the final seed structure and its composition in *Arabidopsis* (Palovaara et al. 2013), *Medicago truncatula* (Gallardo et al. 2007) *Brassica napus* (Li et al. 2005), rice (Furutani et al. 2006), barley (Watson and Henry 2005), and soybean (Collakova et al. 2013). To understand the complete picture of seed in other crops, available information in the public domain may be utilized for data integration to explore the complexity of seed biology at molecular level.

Predictive seed systems biology

It is also known as Bottom-up approach that deals with the molecular mechanism and function of different components present in biological/seed systems via the construction of the mathematical model and formulating hypothesis using the detailed information generated from top-down approaches. The main steps in bottom-up approach are data collection from different resources, pathway modeling, network construction and perturbation analysis through simulation by systems biology tools to predict the behavior of all the components present in systems (Kumar et al. 2015a, b; Pathak et al. 2013; Gupta and Misra 2016).

Most of the available models of biological networks are generally based on transcript accumulation data, which are not able to give a complete picture of cellular metabolism (Gurwitz 2014). Therefore, combined analysis of omics datasets is necessary (Nadeau et al. 1995), and software is being developed to address the difficulty of these “big data” analyses (Hur et al. 2013). Recent studies based on systems approaches clearly demonstrated the integration of metabolomics, transcriptomics and metabolic flux analyses for better understanding of metabolic and regulatory programs that control the seed composition in soybean (Li et al. 2015). The targeted statistical methods, which combined with the Met-Net systems biology platform (Wurtele et al. 2007; Sucaet et al. 2012), were used for analyses of data for development of hypotheses and predictive model in relation to the organization and regulation of the metabolic networks. In addition, 37,593 *Glycine max* probes were annotated via the sequence homology with *Arabidopsis* as model systems to explore the biology of soybean seed (Li et al. 2015).

Intermediate approach in seed biology

To date, understanding the complexity of biological systems is a challenging task for scientific community because omics based high-throughput experiments determine the

complexity of systems at numerous level such as intercommunicating cell categories, interaction of multiple molecules in active pathway and altered states to produces profiling data that is utilized for filling the gaps between top-down and bottom-up approaches. Intermediate approaches are used for integration of various bio-molecules in a structured way from cell to systems level that can quantify the biological systems for filling the unknown information (Butcher et al. 2004; Gupta and Misra 2016).

Tools and databases for seed systems biology

Development of databases and tools for analysis of biological systems is most demanding field today because a team of researchers from various fields such as biological sciences, physical sciences, chemical sciences, mathematical and statistical sciences, and computational sciences having strong backgrounds in said area are required to develop efficient and user-friendly tools and databases with efficient algorithms. Here, we describe some important computational and systems biology tools and databases that should be utilized for deciphering the complexity of seed biology at molecular to systems level for characterization of important genes/protein and act as a functional foods for the development of nutraceuticals and other purposes (Table 1).

Application and expected outcomes

The large-scale sequencing of crop plant genomes, facilitated by hi-throughput omics technologies, data integrations, modeling, visualization of important genes/proteins and simulation analysis approaches has provided novel opportunities, in filling the gap of seed systems and determining its biology through investigating important components present in seed with structural and functional information at molecular to cellular level. Systems biology approaches will be very helpful, when we use diverse genomic, transcriptomic and omics information of seed for analysis of seed systems through integrative and predictive systems biology approaches to develop seed as a model systems with complete information of all the components found in seed and that information can be utilized to develop sustainable seed, nutraceutical product and it can also help in food security in the future (Kumar et al. 2015a, b).

Conclusions

Useful and precise techniques of molecular and systems biology to study the genomics, transcriptomics, proteomics, metabolomics and other omics data of seeds, their integration, and annotation for identification of key components

Table 1 List of some important tools and databases of computational and systems biology

Softwares/databases	Application	Availability
R and bioconductor	It is a well known software packages for analysis of next generation sequencing data, functional genomics, network biology and other applications for deciphering the complexity of biological systems	https://www.r-project.org/ , https://www.bioconductor.org/
CellDesigner	CellDesigner is a program used for pathway modeling, visualization and simulation analysis of biological systems	http://www.celldesigner.org/
Cytoscape	It is highly cited systems biology software for integration and modeling of high-throughput data, visualization and analysis of biological networks	http://www.cytoscape.org/
BiologicalNetworks	It is a tool for integration and modeling of multi-scale data and other applications in network systems biology	http://biologicalnetworks.net/Software/index1.php
Matlab	Matlab is well known tool for modeling and simulation of biological systems	http://in.mathworks.com/products/matlab/?requestedDomain=www.mathworks.com
CLC Genomic Workbench	CLC Genomics Workbench is software to analyze and visualize the high-throughput omics data	http://www.clcbio.com/
PMR (Plant/Eukaryotic and Microbial Systems Resource)	It is database having metabolomics data of plants and eukaryotic microorganisms	http://metnetdb.org/PMR/
KEGG	KEGG is a widely used well known database resources for understanding the complex functions and utilities of biosystems	http://www.genome.jp/kegg/
GEO	GEO is a functional genomics data repository having gene expression data generated through high-throughput omics platform, These data are reusable for the scientific community	http://www.ncbi.nlm.nih.gov/geo/
ArrayExpress	It is a database of the functional genomics experimentation which contains gene expression data from microarray and other high-throughput sequencing platforms for reuse of researchers	https://www.ebi.ac.uk/arrayexpress/
Plant Metabolic Network	PMN is a database that provides information with respect to metabolic pathway in plants	http://www.plantcyc.org/

associated with nutrient loading during seed development followed by their validation is essential to promote translational research to engineer seed systems, which would provide the platform to explore the secret of biochemical seed factories for understanding its nature to improve the nutritional quality and develop sustainable seeds that can easily grow in any environmental conditions and geographical regions. The seed tissue system is influenced by several genes and transcription factors regarding its development. Nutrient distribution event is controlled by a complex network and this network is being successfully explored by these omics-based tools. System biology has perhaps proceeded towards the best designing of nutritional biology research. From the very fundamental point of view, we have reached a well-established platform to continue the research regarding the seed biology for improving nutrition content of seeds. The model plant system has provided the ideas for broader prospectus and approaches making this easier to predict. Therefore, complete informational study needs to be performed that can open a new horizon of molecular and developmental biology studies assisting the nutrition biology research.

Acknowledgements Authors acknowledged Biotechnology Information System Network (BTISNet), Department of Biotechnology (DBT), Govt. of India, New Delhi for the financial assistance. This work was also supported by Science and Engineering Research Board (SERB) New Delhi, India, in the form of “Fast Track Scheme for Young Scientist” awarded to MS (Grant no. YSS/2015/001278) and SG (Grant no. YSS/2015/00536). Author RKP is supported by fellowship from CSIR, India. Bioinformatics Centre (Sub-DIC) at G. B. Pant University of Agriculture and Technology, Pantnagar, India is also acknowledged for providing computational facilities.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Agarwal P, Kapoor S, Tyagi AK (2011) Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* 33(3):189–202
- Anderson A, Hudson M, Chen W, Zhu T (2003) Identification of nutrient partitioning genes participating in rice grain filling by singular value decomposition (SVD) of genome expression data. *BMC Genom* 4(1):26
- Balestrazzi A, Confalonieri M, Macovei A, Donà M, Carbonera D (2011) Genotoxic stress and DNA repair in plants: emerging functions and tools for improving crop productivity. *Plant Cell Rep* 30(3):287–295
- Bevan MW, Uauy C (2013) Genomics reveals new landscapes for crop improvement. *Genome Biol* 14(6):206
- Bewley JD, Bradford K, Hilhorst H (2012) *Seeds: physiology of development, germination and dormancy*. Springer, Berlin
- Brügger B (2014) Lipidomics: analysis of the lipid composition of cells and subcellular organelles by electrospray ionization mass spectrometry. *Annu Rev Biochem* 83:79–98
- Butcher EC, Berg EL, Kunkel EJ (2004) Systems biology in drug discovery. *Nat Biotechnol* 22(10):1253
- Cai J, Okamoto M, Atieno J, Sutton T, Li Y, Miklavcic SJ (2016) Quantifying the onset and progression of plant senescence by color image analysis for high throughput applications. *PLoS One* 11(6):e0157102
- Collakova E, Aghamirzaie D, Fang Y, Klumas C, Tabataba F, Kaku-manu A, Myers E, Heath LS, Grene R (2013) Metabolic and transcriptional reprogramming in developing soybean (*Glycine max*) embryos. *Metabolites* 3(2):347–372
- Dam S, Thaysen-Andersen M, Stenkjær E, Lorentzen A, Roepstorff P, Packer NH, Stougaard J (2013) Combined *N*-glycome and *N*-glycoproteome analysis of the *Lotus japonicus* seed globulin fraction shows conservation of protein structure and glycosylation in legumes. *J Proteome Res* 12(7):3383–3392
- Dean G, Cao Y, Xiang D, Provart NJ, Ramsay L, Ahad A, White R, Selvaraj G, Datla R, Haughn G (2011) Analysis of gene expression patterns during seed coat development in *Arabidopsis*. *Mol Plant* 4(6):1074–1091
- Dekkers BJ, He H, Hanson J, Willems LA, Jamar DC, Cuff G, Rajjou L, Hilhorst HW, Bentsink L (2016) The *Arabidopsis* DELAY OF GERMINATION 1 gene affects ABSCISIC ACID INSENSITIVE 5 (ABI 5) expression and genetically interacts with ABI 3 during *Arabidopsis* seed development. *Plant J* 85(4):451–465
- Deng ZY, Gong CY, Wang T (2013) Use of proteomics to understand seed development in rice. *Proteomics* 13(12–13):1784–1800
- Diaz-Vivancos P, Faize M, Barba-Espin G, Faize L, Petri C, Hernández JA, Burgos L (2013) Ectopic expression of cytosolic superoxide dismutase and ascorbate peroxidase leads to salt stress tolerance in transgenic plums. *Plant Biotechnol J* 11(8):976–985
- Diggle PK, Abrahamson NJ, Baker RL, Barnes MG, Koontz TL, Lay CR, Medeiros JS, Murgel JL, Shaner MG, Simpson HL, Wu CC (2010) Dynamics of maternal and paternal effects on embryo and seed development in wild radish (*Raphanus sativus*). *Ann Bot* 106(2):309–319
- Dumas C, Rogowsky P (2008) Fertilization and early seed formation. *Comptes Rendus Biol* 331(10):715–725
- Edlich-Muth C, Muraya MM, Altmann T, Selbig J (2016) Phenomic prediction of maize hybrids. *Biosystems* 146:102–109
- Eldakak M, Milad SI, Nawar AI, Rohila JS (2013) Proteomics: a biotechnology tool for crop improvement. *Front Plant Sci* 4:35
- El-Maarouf-Bouteau H, Meimoun P, Job C, Job D, Bailly C (2013) Role of protein and mRNA oxidation in seed dormancy and germination. *Front Plant Sci* 4:77
- Engel R (2009) Development of analytical method for determination of water soluble vitamins in functional food product (Doctoral dissertation. Thesis of Ph. D. Dissertation Corvinus University of Budapest, Faculty of Food Science, Department of Applied Chemistry)
- Fang W, Wang Z, Cui R, Li J, Li Y (2012) Maternal control of seed size by EOD3/CYP78A6 in *Arabidopsis thaliana*. *Plant J* 70(6):929–939
- Figueiredo DD, Köhler C (2014) Signaling events regulating seed coat development. *Biochem Soc Trans* 42:358–363
- Fukushima A, Kusano M (2013) Recent progress in the development of metabolome databases for plant systems biology. *Front Plant Sci* 4:73
- Fukushima A, Kanaya S, Nishida K (2014) Integrated network analysis and effective tools in plant systems biology. *Front Plant Sci* 5:598
- Furutani I, Sukegawa S, Kyoizuka J (2006) Genome-wide analysis of spatial and temporal gene expression in rice panicle development. *Plant J* 46(3):503–511

- Gajardo HA, Wittkop B, Soto-Cerda B, Higgins EE, Parkin IA, Snowdon RJ, Federico ML, Iniguez-Luy FL (2015) Association mapping of seed quality traits in *Brassica napus* L. using GWAS and candidate QTL approaches. *Mol Breed* 35(6):143
- Gallardo K, Firnhaber C, Zuber H, Héricher D, Belghazi M, Henry C, Küster H, Thompson R (2007) A combined proteome and transcriptome analysis of developing medicago truncatula seeds evidence for metabolic specialization of maternal and filial tissues. *Mol Cell Proteom* 6(12):2165–2179
- Gaur VS, Singh US, Kumar A (2011) Transcriptional profiling and in silico analysis of Dof transcription factor gene family for understanding their regulation during seed development of rice *Oryza sativa* L. *Mol Biol Rep* 38(4):2827–2848
- Gehring JL, Delph LF (2006) Effects of reduced source-sink ratio on the cost of reproduction in females of *Silene latifolia*. *Int J Plant Sci* 167(4):843–851
- Grant D, Nelson RT, Cannon SB, Shoemaker RC (2009) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucl Acids Res* 38(suppl_1):D843–D846
- Grebe M, Gadea J, Steinmann T, Kientz M, Rahfeld JU, Salchert K, Koncz C, Jürgensa G (2000) A conserved domain of the Arabidopsis GNOM protein mediates subunit interaction and cyclophilin 5 binding. *Plant Cell* 12(3):343–356
- Gu XY, Turnipseed EB, Foley ME (2008) The qSD12 locus controls offspring tissue-imposed seed dormancy in rice. *Genetics* 179(4):2263–2273
- Gupta MK, Misra K (2016) A holistic approach for integration of biological systems and usage in drug discovery. *Netw Model Anal Health Inform Bioinform* 5(1):4
- Gupta N, Gupta AK, Singh NK, Kumar A (2011) Differential expression of PBF Dof transcription factor in different tissues of three finger millet genotypes differing in seed protein content and color. *Plant Mol Biol Rep* 29(1):69–76
- Gupta M, Bhaskar PB, Sriram S, Wang PH (2017) Integration of omics approaches to understand oil/protein content during seed development in oilseed crops. *Plant Cell Rep* 36(5):637–652
- Gupta S, Pathak RK, Gupta SM, Gaur VS, Singh NK, Kumar A (2018) Identification and molecular characterization of Dof transcription factor gene family preferentially expressed in developing spikes of *Eleusine coracana* L.. *3 Biotech* 8(2):82
- Gurwitz D (2014) From transcriptomics to biological networks. *Drug Dev Res* 75(5):267–270
- Gustin JL, Settles AM (2015) Seed phenomics. In: *Phenomics*. Springer, Berlin, pp 67–82
- Hardtke CS, Berleth T (1998) The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J* 17(5):1405–1411
- He D, Yang P (2013) Proteomics of rice seed germination. *Front Plant Sci* 4:246
- He H, Willems LA, Batushansky A, Fait A, Hanson J, Nijveen H, Hilhorst HW, Bentsink L (2016) Effects of parental temperature and nitrate on seed performance are reflected by partly overlapping genetic and metabolic pathways. *Plant Cell Physiol* 57(3):473–487
- Hennig L, Derkacheva M (2009) Diversity of polycomb group complexes in plants: same rules, different players? *Trends Genet* 25(9):414–423
- Hong M, Hu K, Tian T, Li X, Chen L, Zhang Y, Yi B, Wen J, Ma C, Shen J, Fu T (2017) Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. *Front Plant Sci* 8:1674
- Hori K, Sato K, Takeda K (2007) Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. *Theor Appl Genet* 115(6):869–876
- Horn PJ, Chapman KD (2014) Lipidomics in situ: insights into plant lipid metabolism from high resolution spatial maps of metabolites. *Progress Lip Res* 54:32–52
- Hu D, Tateno H, Hirabayashi J (2015) Lectin engineering, a molecular evolutionary approach to expanding the lectin utilities. *Molecules* 20(5):7637–7656
- Hunt L, Bailey KJ, Gray JE (2010) The signalling peptide EPFL9 is a positive regulator of stomatal development. *New Phytol* 186(3):609–614
- Huo H, Wei S, Bradford KJ (2016) DELAY OF GERMINATION1 (DOG1) regulates both seed dormancy and flowering time through microRNA pathways. *Proc Natl Acad Sci* 113(15):E2199–E2206
- Hur M, Campbell AA, Almeida-de-Macedo M, Li L, Ransom N, Jose A, Crispin M, Nikolau BJ, Wurtele ES (2013) A global approach to analysis and interpretation of metabolic data for plant natural product discovery. *Nat Prod Rep* 30(4):565–583
- 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* 15(1):1
- Imtiaz M, Ogbonnaya FC, Oman J, van Ginkel M (2008) Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. *Genetics* 178(3):1725–1736
- Ishibashi Y, Koda Y, Zheng SH, Yuasa T, Iwaya-Inoue M (2012) Regulation of soybean seed germination through ethylene production in response to reactive oxygen species. *Ann Bot* 111(1):95–102
- Jiang H, Köhler C (2012) Evolution, function, and regulation of genomic imprinting in plant seed development. *J Exp Bot* 63(13):4713–4722
- Jing L, Dombinov V, Shen S, Wu Y, Yang L, Wang Y, Frei M (2016) Physiological and genotype-specific factors associated with grain quality changes in rice exposed to high ozone. *Environ Pollut* 210:397–408
- Jofuku KD, Den Boer BG, Van Montagu M, Okamoto JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6(9):1211–1225
- Kazmi RH, Willems LA, Joosen RV, Khan N, Ligterink W, Hilhorst HW (2013) Metabolomic analysis of tomato seed germination. *Metabolomics* 13(12):145
- Kiegle EA, Garden A, Lacchini E, Kater MM (2018) A genomic view of alternative splicing of long non-coding RNAs during rice seed development reveals extensive splicing and lncRNA gene families. *Front Plant Sci* 9:115
- Kigel J (ed) (1995) Seed development and germination, vol 41. CRC Press, Boca Raton, pp 22–29 (ISBN 9780824792299)
- Koller A, Washburn MP, Lange BM, Andon NL, Deciu C, Haynes PA, Hays L, Schieltz D, Ulaszek R, Wei J, Wolters D (2002) Proteomic survey of metabolic pathways in rice. *Proc Natl Acad Sci* 99(18):11969–11974
- Körber 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* 7:386
- Kranner I, Kastberger G, Hartbauer M, Pritchard HW (2010) Noninvasive diagnosis of seed viability using infrared thermography. *Proc Natl Acad Sci* 107(8):3912–3917
- Kumar R, Taware R, Gaur VS, Guru SK, Kumar A (2009) Influence of nitrogen on the expression of TaDof1 transcription factor in wheat and its relationship with photo synthetic and ammonium assimilating efficiency. *Mol Biol Rep* 36(8):2209
- Kumar A, Gaur VS, Goel A, Gupta AK (2015a) De novo assembly and characterization of developing spikes transcriptome of finger millet (*Eleusine coracana*): a minor crop having nutraceutical properties. *Plant Mol Biol Rep* 33(4):905–922

- Kumar A, Pathak RK, Gupta SM, Gaur VS, Pandey D (2015b) Systems biology for smart crops and agricultural innovation: filling the gaps between genotype and phenotype for complex traits linked with robust agricultural productivity and sustainability. *Omics J Integr Biol* 19(10):581–601
- Lariguet P, Ranocha P, De Meyer M, Barbier O, Penel C, Dunand C (2013) Identification of a hydrogen peroxide signalling pathway in the control of light-dependent germination in *Arabidopsis*. *Planta* 238(2):381–395
- Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci* 107(18):8063–8070
- Li F, Wu X, Tsang E, Cutler AJ (2005) Transcriptional profiling of imbibed *Brassica napus* seed. *Genomics* 86(6):718–730
- Li Y, Zheng L, Corke F, Smith C, Bevan MW (2008) Control of final seed and organ size by the DA1 gene family in *Arabidopsis thaliana*. *Genes Dev* 22(10):1331–1336
- Li J, Nie X, Tan JL, Berger F (2013) Integration of epigenetic and genetic controls of seed size by cytokinin in *Arabidopsis*. *Proc Natl Acad Sci* 110(38):15479–15484
- Li L, Hur M, Lee JY, Zhou W, Song Z, Ransom N, Demirkale CY, Nettleton D, Westgate M, Arendsee Z, Iyer V (2015) A systems biology approach toward understanding seed composition in soybean. *In BMC Genom* 16(3):S9 (**BioMed Central**)
- Locascio A, Roig-Villanova I, Bernardi J, Varotto S (2014) Current perspectives on the hormonal control of seed development in *Arabidopsis* and maize: a focus on auxin. *Front Plant Sci* 5:412
- Lohe AR, Chaudhury A (2002) Genetic and epigenetic processes in seed development. *Curr Opin Plant Biol* 5(1):19–25
- Long J, Barton MK (2000) Initiation of axillary and floral meristems in *Arabidopsis*. *Dev Biol* 218(2):341–353
- Lu L, Tian S, Liao H, Zhang J, Yang X, Labavitch JM, Chen W (2013) Analysis of metal element distributions in rice (*Oryza sativa* L.) seeds and relocation during germination based on X-ray fluorescence imaging of Zn, Fe, K, Ca, and Mn. *PLoS One* 8(2):e57360
- Macovei A, Pagano A, Leonetti P, Carbonera D, Balestrazzi A, Araújo SS (2017) Systems biology and genome-wide approaches to unveil the molecular players involved in the pre-germinative metabolism: implications on seed technology traits. *Plant Cell Rep* 36(5):669–688
- Miernyk JA, Hajdich M (2011) Seed proteomics. *J Proteom* 74(4):389–400
- Nadeau J, Lee E, Williams A, Oneill S (1995) Isolation of the *arabidopsis* homologs of 2 genes expressed in *phalaenopsis* ovules—a homeobox gene and a gene of unknown function. *Plant Physiol* 108:117–117
- Navarro PJ, Pérez F, Weiss J, Egea-Cortines M (2016) Machine learning and computer vision system for phenotype data acquisition and analysis in plants. *Sensors* 16(5):641
- Nordine MD, Bartel DP (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev* 24(23):2678–2692
- Nonogaki H (2017) Seed biology updates—highlights and new discoveries in seed dormancy and germination research. *Front Plant Sci* 8:524
- Ogbonnaya FC, Imtiaz M, Ye G, Hearnden PR, Hernandez E, Eastwood RF, Van Ginkel M, Shorter SC, Winchester JM (2008) Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. *Theor Appl Genet* 116(7):891–902
- Ohnishi T, Sekine D, Kinoshita T (2014) Genomic imprinting in plants: what makes the functions of paternal and maternal genes different in endosperm formation? In: *Advances in genetics*, vol 86. Academic Press, London, pp 1–25
- Palovaara J, Saiga S, Weijers D (2013) Transcriptomics approaches in the early *Arabidopsis* embryo. *Trends Plant Sci* 18(9):514–521
- Pathak RK, Taj G, Pandey D, Arora S, Kumar A (2013) Modeling of the MAPK machinery activation in response to various abiotic and biotic stresses in plants by a system biology approach. *Bioinformation* 9(9):443
- Pathak RK, Baunthiyal M, Pandey N, Pandey D, Kumar A (2017) Modeling of the jasmonate signaling pathway in *Arabidopsis thaliana* with respect to pathophysiology of *Alternaria* blight in Brassica. *Sci Rep* 7(1):16790
- Pluskota WE, Martínez-Andújar C, Martin RC, Nonogaki H (2011) Non coding RNAs in plants. In: Erdmann VA, Barciszewski J, Pluskota WE, Martínez-Andújar C, Martin RC, Nonogaki H (eds) *MicroRNA function in seed biology*. Springer, Berlin
- Prasad K, Zhang X, Tobón E, Ambrose BA (2010) The *Arabidopsis* B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. *Plant J* 62(2):203–214
- Raghavan V (2000) *Developmental biology of flowering plants*. Springer, New York
- Rahaman M, Chen D, Gillani Z, Klukas C, Chen M (2015) Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front Plant Sci* 10:619
- Rajendran SR, Yau YY, Pandey D, Kumar A (2015) CRISPR-Cas9 based genome engineering: opportunities in agri-food-nutrition and healthcare. *Omics J Integr Biol* 19(5):261–275
- Ramalingam A, Kudapa H, Pazhamala LT, Weckwerth W, Varshney RK (2015) Proteomics and metabolomics: two emerging areas for legume improvement. *Front Plant Sci* 24:6:1116
- Rangan P, Furtado A, Henry RJ (2017) The transcriptome of the developing grain: a resource for understanding seed development and the molecular control of the functional and nutritional properties of wheat. *BMC Genom* 18(1):766
- Rodrigues AS, Miguel CM (2017) The pivotal role of small non-coding RNAs in the regulation of seed development. *Plant Cell Rep* 36(5):653–667
- Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* 1:tpc-113
- Salekdeh GH, Komatsu S (2007) Crop proteomics: aim at sustainable agriculture of tomorrow. *Proteomics* 7(16):2976–2996
- Sato K, Yamane M, Yamaji N, Kanamori H, Tagiri A, Schwerdt JG, Fincher GB, Matsumoto T, Takeda K, Komatsuda T (2016) Alanine aminotransferase controls seed dormancy in barley. *Nat Commun* 7:11625
- Schrack K, Mayer U, Horrichs A, Kuhnt C, Bellini C, Dangel J, Schmidt J, Jürgens G (2000) FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis embryogenesis*. *Genes Dev* 14(12):1471–1484
- Souza AL, Lopez-Losano JL, Angelo PC, Souza-Neto JN, Cordeiro IB, Astolfi-Filho S, Andrade EV (2017) A proteomic approach to guarana seed and pericarp maturation. *Genet Mol Res* 16(3)
- Sreenivasulu N (2017) Systems biology of seeds: deciphering the molecular mechanisms of seed storage, dormancy and onset of germination. *Plant Cell Rep* 36:633–635
- Sreenivasulu N, Wobus U (2013) Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annu Rev Plant Biol* 64:189–217
- Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci* 98(20):11806–11811
- Sucaet Y, Wang Y, Li J, Wurtele ES (2012) MetNet Online: a novel integrated resource for plant systems biology. *BMC Bioinform* 13(1):267
- Sudarshan GP, Kulkarni M, Akhova L, Ashe P, Shaterian H, Cloutier S, Rowland G, Wei Y, Selvaraj G (2017) QTL mapping and

- molecular characterization of the classical D locus controlling seed and flower color in *Linum usitatissimum* (flax). *Sci Rep* 7(1):15751
- Tang G, Galili G, Zhuang X (2007) RNAi and microRNA: breakthrough technologies for the improvement of plant nutritional value and metabolic engineering. *Metabolomics* 3(3):357–369
- Tao Z, Shen L, Gu X, Wang Y, Yu H, He Y (2017) Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature* 551(7678):124
- Venglat P, Xiang D, Wang E, Datla R (2014) Genomics of seed development: challenges and opportunities for genetic improvement of seed traits in crop plants. *Biocatal Agric Biotechnol* 3(1):24–30
- Wan JM, Cao YJ, Wang CM, Ikehashi H (2005) Quantitative trait loci associated with seed dormancy in rice. *Crop Sci* 45(2):712–716
- Wang H, Schauer N, Usadel B, Frasse P, Zouine M, Hernould M, Latché A, Pech JC, Fernie AR, Bouzayen M (2009) Regulatory features underlying pollination-dependent and -independent tomato fruit set revealed by transcript and primary metabolite profiling. *Plant Cell* 21(5):1428–1452
- Wang HL, Tian CY, Wang L (2017) Germination of dimorphic seeds of *Suaeda aralocaspica* in response to light and salinity conditions during and after cold stratification. *PeerJ* 5:e3671
- Watson L, Henry RJ (2005) Microarray analysis of gene expression in germinating barley embryos (*Hordeum vulgare* L.). *Funct Integr Genom* 5(3):155–162
- Westoby M, Falster DS, Moles AT, Vesk PA, Wright IJ (2002) Plant ecological strategies: some leading dimensions of variation between species. *Annu Rev Ecol Syst* 33(1):125–159
- Willmann MR, Mehalick AJ, Packer RL, Jenik PD (2011) MicroRNAs regulate the timing of embryo maturation in Arabidopsis. *Plant Physiol* 1:pp–110
- Wurtele ES, Li L, Berleant D, Cook D, Dickerson JA, Ding J, Hofmann H, Lawrence M, Lee EK, Li J, Mentzen W (2007) Metnet: systems biology tools for arabidopsis. In: *Concepts in plant metabolomics*. Springer, Dordrecht, pp 145–157
- Xie Q, Mayes S, Sparkes DL (2015) Cernel size, grain filling, and morphology determine individual grain weight in wheat. *J Exp Bot* 66(21):6715–6730
- Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, Yang L, Poethig RS (2016) Developmental functions of miR156-regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes in *Arabidopsis thaliana*. *PLoS Genet* 12(8):e1006263
- Zhai L, Xu L, Wang Y, Huang D, Yu R, Limera C, Gong Y, Liu L (2014) Genome-wide identification of embryogenesis-associated microRNAs in radish (*Raphanus sativus* L.) by high-throughput sequencing. *Plant Mol Biol Rep* 32(4):900–915
- Zhang WH, Zhou Y, Dibley KE, Tyerman SD, Furbank RT, Patrick JW (2007) Nutrient loading of developing seeds. *Funct Plant Biol* 34(4):314–331
- Zhang YC, Yu Y, Wang CY, Li ZY, Liu Q, Xu J, Liao JY, Wang XJ, Qu LH, Chen F, Xin P (2013) Overexpression of microRNA OsmiR397 improves rice yield by increasing grain size and promoting panicle branching. *Nat Biotechnol* 9:848
- Zhou XR, Callahan DL, Shrestha P, Liu Q, Petrie JR, Singh SP (2014) Lipidomic analysis of Arabidopsis seed genetically engineered to contain DHA. *Front Plant Sci* 5:419