

Chapter 3

Omics Platforms: Importance of Twenty-First Century Genome-Enabled Technologies in Seed Developmental Research for Improved Seed Quality and Crop Yield

Ryan D. Syrenne, Weibing Shi, Charles N. Stewart and Joshua S. Yuan

Abstract As the global population increases, so does the demand for caloric intake provided from agricultural crops. Agricultural food production depends predominantly on those species which propagate through seed formation, such as cereals and legumes. The available seed supply of agriculturally important crop varieties with enhanced disease resistance and improved yield is critical for improving food crop production. Traditional techniques used to identify favorable crop characteristics for use in plant breeding are often inadequate in determining specific gene-trait associations. This has resulted in a shift towards integrating plant breeding with new omics technologies. Twenty-first century omics technologies take advantage of many recently released crop genome sequences to investigate gene-function through four disciplines: First, genomics characterizes genome wide expression of DNA; secondly, proteomics studies global protein function and expression; thirdly, transcriptomics is the study of RNA regulation; and fourthly, interactomics is the analysis of complex protein-protein interactions. Examining gene-function of important seed characteristics, through use of omics technology, could reveal critical components that could be exploited for improving seed quality and yield.

Keywords Genomics · Interactomics · Improvement · Proteomics · Seed · Transcriptomics

J. S. Yuan (✉) · R. D. Syrenne · Weibing Shi
Department of Plant Pathology and Microbiology, Texas A & M University,
2123 TAMU, College Station, TX 77843, USA
e-mail: syuan@tamu.edu

R. D. Syrenne · Weibing Shi
Institute for Plant Genomics and Biotechnology, Texas A & M University, 2123 TAMU,
College Station, TX 77843, USA

C. N. Stewart
Department of Plant Science, University of Tennessee, Knoxville, TN 37996, USA

3.1 Introduction

As the global population increases—with an estimated 2 billion more people by 2050 to total 9 billion—so does the demand for caloric intake provided from agricultural crops (Expert Meeting on How to Feed the World et al. 2009). World food production is largely dependent on plant species that propagate through seed formation, such as cereals and legumes, and consequently must rise to meet higher global demand (Borlaug 1983). Advancement in scientific methods and tools has necessitated the need for new technological systems to tackle computational based annotation and comparative genome annotation for the rapid evolution in high-throughput DNA sequencing technologies (Zhang et al. 2010b). As more genome sequences are released, it becomes increasingly important to develop technologies that can rapidly investigate gene function in crop species that correspond to important plant physiological characteristics (Springer and Jackson 2010). Increasingly, application of state-of-the-art omics technologies (i.e., genomics, proteomics, transcriptomics, and metabolomics) are becoming available that hold the promise to detect tissue-specific changes with increasing sensitivity. These methods permit the simultaneous analysis of thousands of genes, proteins or metabolites (Kroeger 2006). This is more relevant than ever, potentially improving overall human wellness through several avenues.

Development of a modern seed industry system is important for providing farmers with consistent, high quality seeds in order to provide optimum crop yield. Coupled with new genome insight, integration of new omics technologies with traditional breeding strategies are vital for providing plant breeders with tools required for commercial criteria of seed production (Flavell 2010). A critical component for improved food crop development and production, especially in developing nations, is the available supply of seed for agriculturally important crop varieties with enhanced disease resistance and improved yield and quality. Traditional trait mapping of favorable crop characteristics used to efficiently breed plants with alleles of interest is often inadequate for precisely identifying favorable gene-trait associations (Flavell 2010). This has resulted in a paradigm shift towards incorporating twenty-first century omics technologies with traditional plant breeding to locate specific gene functions that are important for seed improvement. Providing developing nations with the capacity to produce rapidly improved seed through integrated omics technologies and plant breeding is vital for establishing a mature seed industry that is accessible by the public sector.

Concomitantly to developing an established modern seed industry, many characteristics of economically important crops could potentially improve through state-of-the-art omics technologies. Examining gene function of certain seedling developmental stages, such as grain filling or embryogenesis, could reveal critical components regulating important metabolic processes that could be exploited for improving seed quality (Thompson et al. 2009). Characteristics important for seed improvement include seed dormancy, pathogen and insect resistance, increased abiotic stress tolerance, and carbohydrate accumulation and biosynthetic pathways. The aim of this chapter is to provide an introduction to the so-called omics technologies developing and applying in twenty-first century, their synergistic relationship

with plant breeding, and to elucidate how these technologies have provided a fundamental understanding of seed physiology for improved seed quality and crop yield.

3.2 Biotechnologies Role in Improving Seed Quality and Development

3.2.1 Plant Breeding

The core supply of improved crop varieties has traditionally arisen through farmers selecting and breeding plants with favorable characteristics and then trading those seeds amongst peers. As a result, farmers were limited to select traits observed in the field and improve upon only the natural genetic variation found in the crop species. Because traditional plant breeding at its heart involves the transfer of large genetic portions of parental plants to offspring, specific gene-function association with observed phenotypes is difficult (Flavell 2010; Langridge and Fleury 2011; Varshney et al. 2005). Cross pollination between favorable and unfavorable traits among agricultural crops is difficult for rural farmers to control and often diminishes hybrid vigor, also known as heterosis, resulting in inconsistent seed quality (Guei et al. 2010). The omics platforms can provide support for seed improvement through two aspects. First, crop improvement is essentially dependent upon the natural genetic variation among the species; and breeding programs aim to extend or increase that amount of genetic variation. New omics technologies allow breeders to characterize allelic content of crop germplasm in detail, increasing the ability to discover novel genes and pathways responsible for certain traits (Flavell 2010; Langridge and Fleury 2011). Second, plant breeding requires selection of traits conferred through increased genetic variation. Intensively bred species such as cereal crops often evaluate a million lines every year to maintain high natural genetic variation (Langridge and Fleury 2011) analysis for selection strategies, in part, because of the large computational-based annotation capability and revolutions in DNA sequencing technologies.

3.2.2 Genetic Engineering

For the past millennia, farmers have selected and crossed crops possessing desired traits to improve yield or provide resilient plants through hybrid vigor. Selective pressure towards desirable traits to generate improved crops fundamentally lies in the transfer of allelic material between parent and offspring. This is indirect genetic engineering at its infancy, and has remained with us since the domestication food crops. As human population grew, demand for food and livestock feed increased and crops high in starch, such as cereals and other grain crops, became immensely important (Wan et al. 2008). Contemporary omics platforms and advances in com-

putational biology and other bioinformatics skills enabled the ability to integrate complex gene expression, crosstalk, and protein function of seeds in different developmental stages (Lonneke and van der Geest 2002). Mutant study analysis of seed gene-function through introduction of recombinant DNA, such as T-DNA, to create knockout or over expressed genes has played a significant role in elucidating complex biochemical and metabolic pathways responsible for the dynamic changes that occur during seed development (Lonneke and van der Geest 2002). A more direct approach to colloquially defined genetic engineering involves the insertion of genes into germplasm outside the scope of sexual crossing. The annotated genome of the model reference *Arabidopsis thaliana* has provided a wealth of information for seed biology (North et al. 2010). Inserting genes with known functions into economically important crops and subsequent screening for favorable phenotypes is a classic reverse genetics approach, using genetic engineering, in developing improved seeds. When complimented with new omics technology, use of recombinant DNA can offer exquisite detail of gene-functions, protein interactions, and metabolic biosynthetic pathways important for developing seeds.

3.2.3 Understanding Fundamental Plant Biology for Efficient Concomitant Development of Biotechnology

Twenty-first century omics technologies have enabled in-depth understanding of seed specific cellular processes by observing the abundance of various biological molecules as a function of environmental conditions and genomic expression (Zhang et al. 2010b). High-throughput analysis of various seed molecular components in model plant species (such as *Arabidopsis*) has provided an invaluable tool for development of improved crop cultivars. Common to all seeds are the storage reserves required for seed germination and survival. Ninety percent of seed DW consists of starch (carbohydrates), oils (triacylglycerols), and SPs, and represent the major components of economic value (Ruuska et al. 2002). Advances in fundamental biology have enhanced our understanding of gene function, metabolic regulatory networks, and protein function for economically important seed characteristics. Traditionally, biotechnology has sought to exploit this basic understanding to provide improved seed development for enhanced crop yield through integration of plant breeding and genetic engineering. Recently, these efforts have led to the generation of several new omics technology platforms, system biology, and synthetic biology strategies to facilitate real world application of biotechnology for improved seed development.

3.2.4 Twenty-First Century Omics Platforms

Seed quality improvement and development of high yielding crop varieties revolves around the integration of high-throughput computational data and functional un-

derstanding of cell-wide biological processes. The term omics is collectively used to describe the comprehensive analysis of all biological components of a given system, and often consists of systems biology approaches to assimilate tremendous amounts of data generated by omics technology (Yuan et al. 2008). Generally, shared features of omics technologies and approaches include; holistic and data driven top-down methodologies, the attempt to understand biological gene expression, metabolism, and protein function as one integrated system, and the generation of large datasets (Zhang et al. 2010b). Advances in molecular biology techniques and technologies have resulted in subsequent generation of copious amounts of data, making possible the study of genome-wide expression of DNA (genomics), global protein function and expression (proteomics), RNA regulation and analysis (transcriptomics), and complex protein-protein interactions (interactomics; Singh and Nagaraj 2006). Related to seed improvement and development, omics technologies have allowed detailed study of many facets governing seed biology. In legumes, omics technologies have been used to dissect gene expression at different levels of resolution (organelle, whole seeds), study seed developmental phenomena such as embryogenesis and germination, and understand how biotic and abiotic stresses influence phenotypic traits (Thompson et al. 2009). The remainder of the chapter will describe the evolution and current application towards seed improvement and development of several omics technologies: genomics, transcriptomics, proteomics, and interactomics.

Genomics Seed development and physiology involves many events occurring in different cellular compartments, such as the embryo, endosperm, and seed coat. Therefore, a mosaic of gene expression programs occurs in parallel among seed tissues during different developmental stages (Le et al. 2007). An organism's genome is defined by all the hereditary information encoded in its DNA. Genomics is the study of how the entire set of genes in a given genome are expressed and regulated. Because genomics technologies rely on the sequence of bases in DNA, sequencing technologies have played a crucial role in unraveling the genome.

The original sequencing technology, termed 'Sanger sequencing' for its inventor, was used to determine nucleotide sequences in DNA. Sanger chemistry uses specifically fluorescence labeled nucleotides, Dideoxy nucleoside triphosphates (ddNTPs), which are incorporated into newly synthesized DNA fragments using an original DNA template and DNA polymerase. This allows automated sequencing machines the capability to read through a DNA template during DNA synthesis to determine the order of nucleotides (Zhang et al. 2011). However, even through a series of technological improvements, Sanger sequencing is unable to read fragments of DNA longer than 1 kb which severely limits this technology for sequencing genomes. This is mainly due to the fact that beyond a certain distance from the sequencing primer, very few products are produced because of chain termination.

In order to sequence longer sections of DNA in attempt to obtain the entire genome, a new technology had to be developed. Shotgun sequencing, a DNA sequencing approach developed during the Human Genome Project, allowed the capacity to sequence larger portions of DNA. Shotgun sequencing strategies involve shearing

genomic DNA into smaller fragments, either enzymatically or mechanically, and cloning into sequencing vectors such as Bacterial Artificial Chromosomes (BAC) where they can be sequenced individually. Cloning large fragments of DNA into BACs is then used to create a genomic library for a set of recombinant clones that contain the entire DNA for a given organism. These libraries with large genomic inserts are important for genome sequencing and have become invaluable genomic tools because of BACs inherent ability to maintain large DNA fragments for easy manipulation (Shizuya et al. 1992).

Next-generation genome sequencing technology, or massively parallel sequencing, uses the core philosophy of shotgun sequencing approaches for generating cost effective and efficient methods of genome sequencing (Zhang et al. 2011). Next-generation sequencing (NGS) technology reads the DNA template along the entire genome by first fragmenting the DNA, and then ligating the DNA fragments to adaptors that are randomly read during DNA synthesis. However, because the read lengths are typically shorter than those produced by Sanger sequencing (50–500 base pairs), increasing coverage is important. Coverage is the total amount of fragmented DNA sequences that overlap within a specific region and is important to accurately assemble the fragments (Zhang et al. 2011). Because of the prohibitive costs associated with sequencing and assembling large eukaryotic genomes, next-generation massively parallel sequencing using only the Illumina Genome Analyzer was first carried out for the giant panda genome (Li et al. 2010a).

Comparative genomics is an approach used in conjunction with the aforementioned technologies to exploit similarities between sequenced model species, such as *Arabidopsis*, and species of interest in order to identify novel genes and infer gene function (Buckley 2007). So far, there are more than 25 plant model organisms were applied for whole genome sequencing, such as beans, rice, grasses, corn, maize, grape, sorghum, banana, wheat, etc. (<http://www.arabidopsis.org/>). The release of these whole genomes sequencing data makes the life much easier for molecular breeding. Because closely related cultivars generally used for crossing material lack sufficient known DNA polymorphisms due to their genetic relatedness, and next-generation sequencing allows the identification of a massive number of DNA polymorphisms, such as nucleotide polymorphisms (SNPs) and insertions-deletions between highly homologous genomes. Most recently, Arai-Kichise et al. (2011) performed whole-genome sequencing of a landrace of japonica rice, Omachi, which is used for sake brewing and is an important source for modern cultivars. In this study, they identified 132,462 SNPs, 16,448 Insertions, and 19,318 deletions between the Omachi and Nipponbare genomes (Arai-Kichise et al. 2011). This is incredible for conventional molecular breeding scientists. There are also numerous whole-genome sequencing works that have been done, and all of these efforts will dramatically reform the situation of molecular breeding and seed development.

Genomics technologies have elucidated many complex physiological and developmental characteristics exhibited during different stages of seed growth. Information gained through genomics studies of seeds have led to various discoveries that could prove useful in applications for improved seed development. Gene expression studies have led to the discovery of genes that contribute to cell wall weakening

during germination, such as expansin, and other genes such as those involved in energy metabolism (Bradford et al. 2000; Lonneke and van der Geest 2002).

Transcriptomics The measure of all mRNA molecules or transcripts produced in a given cell, often called genome-wide expression profiling, is known as transcriptomics (Zhang et al. 2010b). Transcriptomics have allowed the development of analyses that relate the abundance of mRNA molecules, or transcripts, to gene expression and regulation under varying environmental conditions. As a result, transcriptomics is the dynamic link between genomics and proteomics and thus can elaborate on the complex cellular processes responsible for adapting to environmental conditions (Singh and Nagaraj 2006). Seed development is associated with enormous differential gene expression depending on growth stage, setting the importance of transcriptomics in context for seed development and improvement.

Like genomics, transcriptomics has seen a revolution in technological improvements. Microarray technology has been at the forefront for studying simultaneous changes in gene expression and quantification across the entire genome. Microarray technology essentially works through the principle of DNA hybridization. That is, when mRNA molecules are fluorescently labeled and mixed with individual genes of an organism on a microarray plate, single mRNA molecules will hybridize with only one complimentary DNA strand. The amount of mRNA that attaches to its complimentary DNA strand is proportional to the abundance of mRNA transcripts in the sample. In its infancy, microarray technology used only one labeled probe to attach to mRNA molecules, which allowed only the expression of genes under one condition of the sample to be analyzed. Two-color microarray technology, developed by Patrick Brown, allowed the differential quantification and gene expression of two samples (test and control) based on two different labeled probes, green cyanine (Cy)3 and red Cy 5 (Schena et al. 1995).

Affymetrix microarrays, as compared to two-color systems, use one color-labeled fluorescent probe to detect mRNA hybridizations with cRNA. Several advantages lie with the Affymetrix system over traditional two-color microarrays: first, the fabrication process for Affymetrix GeneChips allows for very high density arrays to be produced, and second, because a single sample is hybridized to each array more precision in measurement is achieved by minimizing array-array variability often seen in two-color microarrays (de Reynies et al. 2006; Woo et al. 2004). Also, it was found that single color Affymetrix GeneChips were capable of higher levels of reproducibility compared to two-color microarrays (de Reynies et al. 2006).

Concomitant to the two-color microarray systems for quantifying transcripts in a cell, in fact published in the same article of Science, another technique for measuring gene expression based on mRNA quantification was developed, called Serial Analysis of Gene Expression (SAGE). Two principles define SAGE: first, short nucleotide sequence tags contain enough information to identify individual transcripts, even among thousands; and second, joining sequence tags into a string followed by sequencing in a single clone allows for efficient analysis of transcripts (Velculescu et al. 1995). Quantification and identification of novel expressed genes

can also be accomplished through using SAGE, as well as the analysis of eukaryotic genomes other than the human genome.

Specific hybridizations of probes to microarrays or GeneChips, and counting of tags on DNA fragments used in SAGE are subject to many sources of variation. In tackling the shortcomings of these technologies, Brenner et al. (2000) developed a technology that does not require physical separation of DNA fragments, termed Massively Parallel Signature Sequencing (MPSS). MPSS benefits reside in the ability to handle highly variable mixtures of nucleic acid fragments and subsequent cloning onto a million microbeads (Brenner et al. 2000). Millions of template microbeads are then analyzed and sequenced through flow cytometry using a fluorescent imaging system. Since MPSS does not require separate isolation and processing of templates, it is suited well for gene expression analysis.

Technical advancements in high-throughput mRNA transcript analysis through technologies such as microarray, Affymetrix, and multiple parallel signature sequencing have greatly contributed to our understanding of biological gene function. Recently, the tools developed for massively parallel sequencing have brought about another major breakthrough in the field of transcriptomics, known as RNA sequencing (RNA-seq; Malone and Oliver 2011). While microarray technology was used for genome-wide expression studies during the 1990s, new RNA-seq technology offers several select advantages, making it an attractive alternative. RNA-seq technologies, such as those offered by Illumina, sample the amount of starting transcripts present in the tissue by direct sequencing of cDNA in order to derive information about RNA (Malone and Oliver 2011). Molecular hybridization to a probe for capture and determination of transcript abundance is often irrespective of the actual transcript expression level, resulting in reproducibility errors and limiting the accuracy of gene expression experiments (Marioni et al. 2008). RNA-seq experiments using Illumina are, on the other hand, highly reproducible with little technical variation when compared to traditional gene expression assays (Marioni et al. 2008). Also, since RNA-seq provides direct information about the sequence, it is useful for measuring gene expression for organisms which there is no full genome sequence (Malone and Oliver 2011).

Recently, transcriptomics has become more and more of a hot topic. For instance, Yang and coworkers presented the first comprehensive characterization of the *O. longistaminata* root transcriptome using 454 pyrosequencing, and revealed expressed genes across different plant species, and also identified 15.7 % of unknown ESTs (Yang et al. 2010a). These sequence data are publicly available and will facilitate gene discovery and seed development studies. RNA-seq technology also used to explore the transcriptome of a single plant cell type, the *Arabidopsis* mal meiocyte, detecting the expression of approximately 20,000 genes (Yang et al. 2011b). In this study, Yang et al. (2011b) identified more than 1000 orthologous gene clusters that are also expressed in meiotic cells of mouse and fission yeast. These RNA-seq transcriptome data provide an overview of gene expression in male meiocytes and invaluable information for future functional studies.

Transcriptomics data has provided a valuable resource for molecular studies of many important agricultural crops, such as cereals and legumes. Affymetrix

arrays have proved invaluable for developing the transcriptome and elucidating seed developmental changes for grain development of wheat (Wan et al. 2008). RNA-seq using Illumina has provided a high quality in-depth view of the Soybean transcriptome during seed developmental stages and has potentially identified novel genes previously undetected (Severin et al. 2010). Combined microarrays, Affymetrix fluorescent fluorophores GeneChip's, MPSS, and next-generation RNA-Seq have contributed greatly to our current understanding of the seed transcriptome.

Proteomics Proteins are essential components for organisms since they comprise many of the biological building components, catalytic enzymes for metabolic pathways, and signal transduction elements in the cell (Zhang et al. 2010b). When integrated with previously mentioned omics technologies, proteomics provides powerful insight into referencing protein properties and functions for improved seed development and crop yield (Tyers and Mann 2003). Proteomics analysis is critical in improved seed development for incorporation into modern plant breeding strategies because proteins directly affect phenotypic traits as opposed to gene sequence (Singh and Nagaraj 2006).

The ability to detect and measure protein molecules has traditionally relied on several methods to separate and observe protein structure: two-dimensional PAGE (2-DGE) and mass spectrometry (Zhang et al. 2010b). 2-DGE separates proteins based on two distinct properties: mass and isoelectric point (pI). Basically, the protein is first separated based on its isoelectric point; a property wherein proteins will remain charged at all pH values other than their pI . This is accomplished through use of a pH gradient. After proteins are separated based on isoelectric point, they are washed and treated with sodium dodecyl sulfate (SDS) to denature the protein and yield a linear, negatively charged polypeptide. Separation of the protein is then done by applying an electric current to the gel, where proteins are separated by size similar to traditional gel electrophoresis (for review see, Righetti et al. 2008).

Difference gel electrophoresis (DIGE) is similar to 2-DGE, but involves labeling the amino group of lysines in the polypeptides with fluorescent fluorophores (Cy2, Cy3, and Cy5; Wu et al. 2006). DIGE allows the comparison of proteins for different treated samples on the same gel, thus removing inter-gel variation commonly associated with traditional 2-DGE. Even though 2-DGE has undergone several technical innovations, protein coverage still lacks compared to gel-free proteomics techniques, such as mass spectrometry (MS)-based shotgun proteomics.

MS-based shotgun proteomics approach for protein identification and quantification are quickly becoming the preferred method, namely for their high-throughput platforms and ability to analyze complex samples at high resolution (Webb-Robertson and Cannon 2007). Traditionally, proteomics studies relied on protein separation via 2-DGE followed by identification of excised gel bands using MS. However, this method is limited by the fact that only the most abundant proteins in the sample are analyzed, which leads to an issue of coverage (Tyers and Mann 2003). Gel-free analysis of complex mixtures now dominates the repertoire of analytical techniques

used for protein identification of low-abundance peptides. Liquid chromatography (LC) tandem MS (LC-MS/MS) is one such preferred analytical technique for protein identification.

Multidimensional protein identification technology (MudPIT) is quickly becoming a common approach for protein identification using gel-free based LC-MS/MS methods. Relative protein quantitation using MudPIT technology takes essentially two different approaches: label and label free (Cooper et al. 2010). The most popular stable-isotope labeling uses isotope-coded affinity tags (ICAT) technology for protein quantitation (Wu et al. 2006). Essentially, isobaric tags and prototypical isotope labeling results in a shift of labeled protein mass and allows for comparison between differentially labeled ions, ion peak height, and the resulting mass spectra produced (Cooper et al. 2010). Label-free approaches are also used to determine amounts of peptide in sample mixtures, and are based on the relationship between protein abundance, determined by the mass spectral peak analysis, and sampling statistical analysis (Bridges et al. 2007; Cooper et al. 2010). MudPIT technology first requires site specific enzymatic digestion of protein followed by separation of digested fragments using liquid chromatography, typically with a C18 reverse-phase (RP) column, based on cation exchange (Wu et al. 2006; Bridges et al. 2007). Protein fragments are then subjected to MS/MS for protein identification and quantification.

Right now, proteomics is a very popular technology and platform, and has been widely applied in plant tissue specific studies alone or together with other omics technologies. In previous studies, proteomics approach has been used to study the mechanism of action of grape seed proanthocyanidin extracts on arterial remodeling in diabetic rats (Li et al. 2010b). It also used to identify pathogenesis-related (PR) proteins in rice (Yang et al. 2011c), soybean (Yang et al. 2011a), etc. Some scientists used proteomics technology to identify the stress-related proteins (Lee et al. 2011; Wen et al. 2010; Zhou et al. 2011). Integration of proteomics and transcriptomics, even metabolomics technologies provides deep and new insights into the plant systems biology (Fig. 3.1). For instance, Barros and colleagues applied three profiling technologies to compare the transcriptome, proteome, and metabolome of two transgenic maize lines with the respective control line, and revealed that the environment was shown to play an important effect in the protein, gene expression and metabolite levels of the maize samples (Barros et al. 2010).

In addition, shotgun proteomics technology has also been used to help develop the proteome and understand the metabolic processes of seed filling in soybean. It was found by Agrawal and coworkers, that when using 2-DGE, membrane proteins were under represented in soybean samples when compared to MudPIT technology (Agrawal et al. 2008). However, proteins involved in metabolism, protein transport, and energy were equally represented using both 2-D gel technologies and MudPIT, highlighting the respective strengths of each technology for application in improved seed development (Agrawal et al. 2008).

Interactomics Understanding genome-wide mRNA expression profiling has provided exceptional information regarding gene-function analyses studies, but lacks

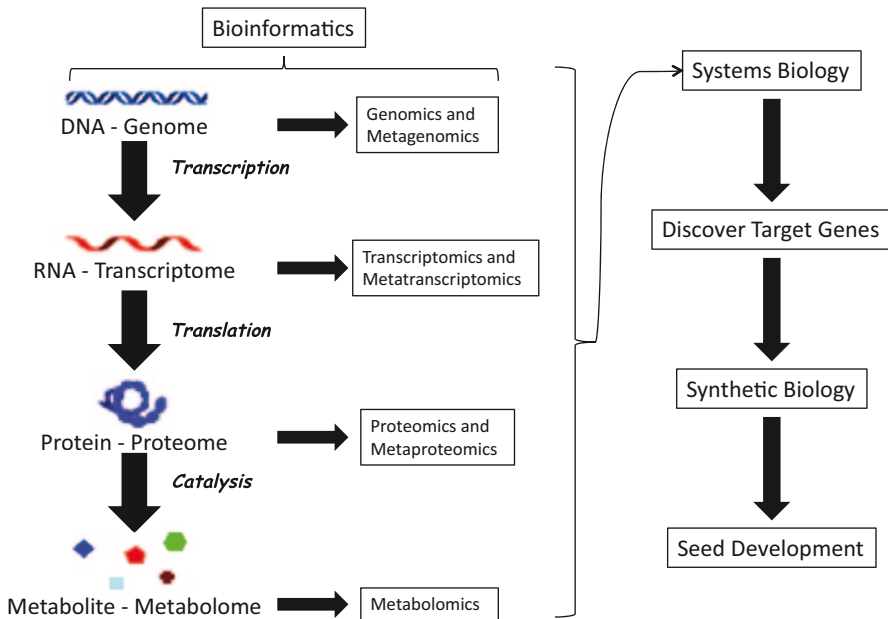


Fig. 3.1 Overview of 21st 'omics' technologies for seed development

the ability to provide information regarding gene product destination, and post-transcriptional modification (PTM) behaviors (Singh and Nagaraj 2006). The large scale detection of aggregated protein complexes in similar functional pathways, coupled with protein-protein interactions is collectively defined as interactomics (Zhang et al. 2010a). Protein-protein interactions are important considerations for understanding terminal gene function of different seed developmental stages and metabolic processes, such as embryogenesis and grain filling. Understanding how proteins interact could provide the fundamental knowledge necessary for improved seed development and crop yield.

The recent advent of the systems biology age and the accumulation of tremendous amounts of data concerning protein-protein interaction networks have resulted in several major technologies. Yeast two hybrid (Y2H) systems and affinity purification MS (AP-MS) have underpinned major technological breakthroughs for understanding protein interaction networks (Zhang et al. 2010a).

Y2H systems, first introduced in 1989, have seen a major evolution in technical improvement, resulting in adaptation for the various needs in studying protein-protein interaction networks (Zhang et al. 2010a). Y2H is a molecular biology technique used to identify association between proteins based on interaction of transcription factor components to express a reporter gene, such as green fluorescent protein (GFP). Fundamentally, Y2H relies on the ability of fragmented transcription factors to function when concatenated under close proximity to activate down-

stream reporter gene expression. The yeast transcription factor GAL4 is often used in Y2H studies because it binds with the UAS promoter and contains two functional domains; the transcription activation domain and the DNA-binding domain (Zhang et al. 2010a). Y2H fuses these two domains with a bait protein, which is usually known, and ‘prey’ protein typically derived from a library of known or unknown proteins, and relies on the interaction with bait and target ‘prey’ proteins to form a functional GAL4 transcription factor to activate UAS-driven expression of the reporter genes.

However, limitations of the Y2H system for identify protein interactions have resulted in incomplete data for developing a comprehensive interactome. For example, Y2H systems have a tendency for increasing high false-positive rates for protein interactions, in particular for membrane proteins and proteins with strong expression signals in the cytoplasm (Zhang et al. 2010a). This is due because Y2H interaction systems are often limited to the nucleus.

AP-MS has proven more successful for identifying protein-protein interactions and components of protein complexes through highlighting several of the major limitation present in Y2H techniques (Zhang et al. 2010a). An intrinsic limitation with Y2H technology is that only two proteins can be examined for interaction, whereas AP-MS is capable of elucidating protein interactions of much larger protein complexes.

AP-MS combine’s affinity purification of bait protein’s labeled with affinity tags, such as his-tag, with MS to identify novel protein complexes (Van Leene et al. 2008; Zhang et al. 2010a). In AP-MS, multi-component protein complexes are isolated from cell’s *via* affinity purification of the labeled tags and then subject to downstream analysis by MS or MS/MS. Several advantages of AP-MS techniques for identifying protein-protein interaction complexes are apparent: first, AP-MS is conducted under native physiological conditions, which represents a more accurate picture of *in vivo* protein binding compared to Y2H; and second, the dynamics of protein interactions can be observed under various conditions (Zhang et al. 2010a).

Bimolecular fluorescent complementation (BiFC) is another molecular biological approach often used to examine peptide interactions. BiFC approaches are based on the idea that when two proteins are close together, the probability of interaction is high. Basically, BiFC uses a split, nonfunctional, fluorescent reporter protein combined to a bait and target protein. When these two nonfunctional fluorescent reporter components are reconstituted, through close proximal interaction of bait and labeled proteins, a functional GFP complex is formed to emit an observable signal. Self-fluorescence of plants generally precludes the use of GFP, and instead, red fluorescent or yellow fluorescent proteins are used as reporters (Zhang et al. 2010a). Although not considered a high-throughput system, and thus a major drawback as a next-generation protein interaction analytical tool, BiFC advantages lie in its high sensitivity and its ability to be used to explore protein interactions in various cellular compartments.

3.3 Concluding Remarks

The twenty-first century omics technologies are an essential component for improved seed development and crop yield. Genomics, transcriptomics, proteomics, and interactomics have provided biologists with comprehensive data concerning global gene function and identification, dynamic changes between gene expression, protein identification and function, and interaction of entire protein networks. When combined, these information technologies present a snapshot of entire cellular functional pathways, such as genome structure and metabolic regulatory networks. Combined, the twenty-first century omics are capable of characterizing a comprehensive picture of seed developmental physiology and morphology, vital for transcending the gap between fundamental biology knowledge and application towards improved seeds important for establishing modern seed industry and plant breeding programs for increased food production.

References

- 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
- Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, Yano M, Wakasa K (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of *Japonica* rice by whole-genome sequencing. *Plant Cell Physiol* 52:274–282
- Barros E, Lezar S, Anttonen MJ, Van Dijk JP, Röhlrig RM, Kok EJ, Engel KH (2010) Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnol J* 8:436–451
- Borlaug NE (1983) Contributions of conventional plant breeding to food production. *Science* 219:689–693
- Bradford KJ, Chen F, Cooley MB, Dahal P, Downie B, Fukunaga KK, Gee OH, Gurusinghe S, Mella RA, Nonogaki H, Wu CT, Yang H, Yim KO (2000) Gene expression prior to radicle emergence in imbibed tomato seeds. In: seed biology: advances and applications. Proceedings of the sixth international workshop on seeds, Merida, Mexico, 1999, Black M, Bradford KJ, Vazquez-Ramos J (eds), CABI, CABI International, Wallingford, pp 231–251
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18:630–634
- Bridges SM, Magee GB, Wang N, Williams WP, Burgess SC, Nanduri B (2007) ProtQuant: a tool for the label-free quantification of MudPIT proteomics data. *BMC Bioinformatics* 8(Suppl 7):S24
- Buckley BA (2007) Comparative environmental genomics in non-model species: using heterologous hybridization to DNA-based microarrays. *J Exp Biol* 210:1602–1606
- Cooper B, Feng J, Garrett WM (2010) Relative, label-free protein quantitation: spectral counting error statistics from nine replicate MudPIT samples. *J Am Soc Mass Spectrom* 21:1534–1546
- de Reynies A, Geromin D, Cayuela JM, Petel F, Dessen P, Sigaux F, Rickman DS (2006) Comparison of the latest commercial short and long oligonucleotide microarray technologies. *BMC Genomics* 7:51

- Expert Meeting on How to Feed the World, i., Food and Agriculture Organization of the United Nations (2009) Proceedings of the expert meeting on how to feed the world in 2050 (electronic resource), FAO Headquarters, Rome, 24–26 June 2009
- Flavell R (2010) From genomics to crop breeding. *Nat Biotechnol* 28:144–145
- Guei RG, Food and Agriculture Organization of the United Nations (2010) Promoting the growth and development of smallholder seed enterprises for food security crops: case studies from Brazil, Côte d'Ivoire and India/coordinated by Robert G. Guei. Rome, Food and Agriculture Organization of the United Nations
- Kroeger M (2006) How omics technologies can contribute to the '3R' principles by introducing new strategies in animal testing. *Trends Biotechnol* 24:343–346
- Langridge P, Fleury D (2011) Making the most of 'omics' for crop breeding. *Trends Biotechnol* 29:33–40
- Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB (2007) Using genomics to study legume seed development. *Plant Physiol* 144:562–574
- Lee DG, Park KW, An JY, Sohn YG, Ha JK, Kim HY, Bae DW, Lee KH, Kang NJ, Lee BH, Kang KY, Lee JJ (2011) Proteomics analysis of salt-induced leaf proteins in two rice germplasms with different salt sensitivity. *Can J Plant Sci* 91:337–349
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, Zhang Z, Zhang Y, Wang W, Li J, Wei F, Li H, Jian M, Li J, Zhang Z, Nielsen R, Li D, Gu W, Yang Z, Xuan Z, Ryder OA, Leung FCC, Zhou Y, Cao J, Sun X, Fu Y, Fang X, Guo X, Wang B, Hou R, Shen F, Mu B, Ni P, Lin R, Qian W, Wang G, Yu C, Nie W, Wang J, Wu Z, Liang H, Min J, Wu Q, Cheng S, Ruan J, Wang M, Shi Z, Wen M, Liu B, Ren X, Zheng H, Dong D, Cook K, Shan G, Zhang H, Kosiol C, Xie X, Lu Z, Zheng H, Li Y, Steiner CC, Lam TTY, Lin S, Zhang Q, Li G, Tian J, Gong T, Liu H, Zhang D, Fang L, Ye C, Zhang J, Hu W, Xu A, Ren Y, Zhang G, Bruford MW, Li Q, Ma L, Guo Y, An N, Hu Y, Zheng Y, Shi Y, Li Z, Liu Q, Chen Y, Zhao J, Qu N, Zhao S, Tian F, Wang X, Wang H, Xu L, Liu X, Vinar T, Wang Y, Lam TW, Yiu SM, Liu S, Zhang H, Li D, Huang Y, Wang X, Yang G, Jiang Z, Wang J, Qin N, Li L, Li J, Bolund L, Kristiansen K, Wong GKS, Olson M, Zhang X, Li S, Yang H, Wang J, Wang J (2010a) The sequence and *de novo* assembly of the giant panda genome. *Nature* 463:311–317
- Li XL, Li BY, Gao HQ, Cheng M, Xu L, Li XH, Zhang WD, Hu JW (2010b) Proteomics approach to study the mechanism of action of grape seed proanthocyanidin extracts on arterial remodeling in diabetic rats. *Int J Mol Med* 25:237–248
- Lonneke A, van der Geest HM (2002) Seed genomics: germinating opportunities. *Seed Sci Res* 12:145–153
- Malone J, Oliver B (2011) Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol* 9:34
- 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
- North H, Baud S, Debeaujon I, Dubos C, Dubreucq B, Grappin P, Jullien M, Lepiniec L, Marion-Poll A, Miquel M, Rajjou L, Routaboul JM, Caboche M (2010) *Arabidopsis* seed secrets unravelled after a decade of genetic and omics-driven research. *Plant J* 61:971–981
- Righetti PG, Antonioli P, Simò C, Citterio A (2008) Gel-based proteomics. In: Agrawal GK, Rakwal R (eds) *Plant proteomics: technologies, strategies, and applications*. Wiley, Hoboken, pp 11–30
- Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell* 14:1191–1206
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science* 270:467–470
- Severin A, Woody J, Bolon YT, Joseph B, Diers B, Farmer A, Muehlbauer G, Nelson R, Grant D, Specht J, Graham M, Cannon S, May G, Vance C, Shoemaker R (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol* 10:160
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A* 89:8794–8797

- Singh OV, Nagaraj NS (2006) Transcriptomics, proteomics and interactomics: unique approaches to track the insights of bioremediation. *Brief Funct Genomics Proteomics* 4:355–362
- Springer NM, Jackson SA (2010) Realizing the potential of genomics for crop improvement. *Brief Funct Genomics* 9:93–94
- Thompson R, Burstin J, Gallardo K (2009) Post-genomics studies of developmental processes in legume seeds. *Plant Physiol* 151:1023–1029
- Tyers M, Mann M (2003) From genomics to proteomics. *Nature* 422:193–197
- Van Leene J, Witters E, Inzé D, De Jaeger G (2008) Boosting tandem affinity purification of plant protein complexes. *Trends Plant Sci* 13:517–520
- Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci* 10:621–630
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. *Science* 270:484–487
- Wan YF, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, Gooding MJ, Mills C, Edwards KJ, Shewry PR, Mitchell RAC (2008) Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics* 9:121
- Webb-Robertson BJM, Cannon WR (2007) Current trends in computational inference from mass spectrometry-based proteomics. *Brief Bioinformatics* 8:304–317
- Wen FP, Zhang ZH, Bai T, Xu Q, Pan YH (2010) Proteomics reveals the effects of gibberellic acid GA3 on salt-stressed rice (*Oryza sativa* L.) shoots. *Plant Sci* 178:170–175
- Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggart J, Churchill G (2004) A Comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. *J Biomol Tech* 15:276–284
- Wu WW, Wang G, Baek SJ, Shen RF (2006) Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J Proteome Res* 5:651–658
- Yang H, Huang YP, Zhi HJ, Yu DY (2011a) Proteomics-based analysis of novel genes involved in response toward soybean mosaic virus infection. *Mol Biol Rep* 38:511–521
- Yang H, Lu P, Wang Y, Ma H (2011b) The transcriptome landscape of *Arabidopsis* male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. *Plant J* 65:503–516
- Yang L, Su N, Wu M, Wang C, Hu H (2011c) Proteomics to identify pathogenesis-related proteins in rice roots under water deficit. *Biologia* 66:477–483
- Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart CN Jr (2008) Plant systems biology comes of age. *Trends Plant Sci* 13:165–171
- Zhang Y, Gao P, Yuan JS (2010a) Plant protein-protein interaction network and interactome. *Curr Genomics* 11:40–46
- Zhang W, Li F, Nie L. (2010b) Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiol* 156:287–301
- Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. *J Genet Genomics* 38:95–109
- Zhou L, Bokhari SA, Dong CJ, Liu JY (2011) Comparative proteomics analysis of the root apoplasts of rice seedlings in response to hydrogen peroxide. *PLoS One* 6:e16723