Journal of Pharmacognosy and Phytochemistry
Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2017; 6(5): 941-950
Received: 19-07-2017
Accepted: 20-08-2017

Ashok
Phd Scholar, Department of Seed Science and technology, UAS, Raichur, Karnataka, India

Shakuntala NM
Professor and Head, Department of Seed Science and technology, UAS, Raichur, Karnataka, India

Basave gowda
Professor (SST) and Special Officer (Seeds), Seed Unit, UAS, Raichur, Karnataka, India

Correspondence
Ashok
Phd Scholar, Department of Seed Science and technology, UAS, Raichur, Karnataka, India

Omics in seed development: A key factor for improvement of seed yield and quality

Ashok, Shakuntala NM and Basave gowda

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 favourable 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: interactomics, proteomics, seed development, transcriptomics, metabolomics

1. Introduction
'OMIC' is a field of study in biology, refers to the collective technologies used to explore the roles, relationships and actions of the various types of molecules that make up the cells of an organism. They are aimed primarily at the universal detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample in a non-targeted and non-biased manner. This can also be referred to as high-dimensional biology; the integration of these techniques is called systems biology (Saito and Matsuda, 2010) [14].

As the global population increases, the demand for caloric intake provided from agricultural crops. 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.

A seed is a small embryonic plant enclosed in a covering called the seed coat, usually with some stored food. Seed development is a dynamic process coordinated by the three distinct organs of a seed: Embryo, Endosperm and Seed coat (Radchuk and Borisjuk 2006) [12]. The diploid embryo and triploid endosperm represent the filial lineages, whereas the testa represents the diploid maternal generation. The endosperm initially develops as a syncytium then cellularises to encircle the developing embryo. The endosperm either remains to feed the growing embryo and to store food reserves in mature seeds. The seed coat, which develops by differentiation of two ovule integuments, not only protects the developing embryo and endosperm but also has other roles such as metabolic control of seed development and dormancy, metabolism of nutrients from parent plant and disease resistance (Weber *et al.* 2005) [16].

Omics: Importance in seed development
- This study will helps to know the genes responsible for differentiation of zygote
- Initiation for the embryo and development
- Development of endosperm

- Metabolism and Storage of food reserves
- Seed size, coat and colour development

a. Genomics

Genomics is a discipline in genetics that applies recombinant DNA, DNA sequencing methods and bioinformatics to sequence, assemble and analyze the function and structure of genomes. Genomics is the new science that deals with the discovery and noting of all the sequences in the entire genome of a particular organism. The genome can be defined as the complete set of genes inside a cell. Genomics is therefore, the study of the genetic make-up of organisms.

Genomics is an entry point for looking at the other 'omics' sciences. The information in the genes of an organism, its genotype, is largely responsible for the final physical makeup of the organism, referred to as the "phenotype". However, the environment also has some influence on the phenotype. DNA in the genome is only one aspect of the complex mechanism that keeps an organism running - so decoding the DNA is one step towards understanding the process. However, by itself, it does not specify everything that happens within the organism.

The basic flow of genetic information in a cell is as follows. The DNA is transcribed or copied into a form known as "RNA". The complete set of RNA (also known as its transcriptome) is subject to some editing (cutting and pasting) to become messenger-RNA, which carries information to the ribosome, the protein factory of the cell, which then translates the message into protein.

Types of genomics

1. **Structural Genomics**: Sequencing the genome to get the content & organization of genome
2. **Functional Genomics:** To understand the function of information in genome
3. **Comparative Genomics:** Comparing the genomes of different organisms

Genomics applied to agriculture

- Sequencing of crop-plant genomes (whole genome sequence)
- Gene discovery for useful traits (find trait controlling genes)
- What DNA information available –what is the practical goal
- Genome-wide regulatory networks to improve traits
- The field of genomics deals with the DNA sequence, organization, function, and evolution of genomes.

DNA sequencing methods

1. Basic DNA sequencing

- Sanger method
- Maxam gilbert method

2. Advanced DNA sequencing

- Short gun sequencing

3. Next generation sequencing

- Solid sequencing
- Ilumina sequencing
- Pyrosequencing

Importance of genomics in seed development as given below as follows:

- It is now possible to identify most of the genes responsible for guiding seed development in every cell, tissue and organ throughout the seed lifecycle.
- Genomics-based research will provide the necessary tools to improve seeds: seeds with improved nutritional value, that can endure adverse environmental conditions, or one that can withstand biological attack.
- Many mRNAs coded by the specific genes that are constitute the top 50% most abundant mRNAs in one or two cell embryos.
- Zygotic genome is activated immediately after fertilization and plays a major regulatory role during early embryo genesis (Scott *et al*., 1998).
- Many genes related to food (*Sucrose-proton symporter 2*) and water transport (*Tonoplast intrinsic proteins 1*) are found in chalazal seed coat.
- Up to 80% of the genes in angiosperm genomes are transcribed during seed development, but only a small number (~3%) are expressed exclusively in the seed or in specific seed tissues/cell types.

1. The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp by Furukawa *et al.* (2006)

Different colors, such as purple, brown, red and white, occur in the pericarp of rice. Here, two genes affecting proanthocyanidin synthesis in red- and brown-colored rice were elucidated. Genetic segregation analysis suggested that the Rd and A loci are identical, and both encode dihydroflavonol-4-reductase (DFR). The introduction of the *DFR* gene into an *Rcrd* mutant resulted in red-colored rice, which was brown in the original mutant, demonstrating that the Rd locus encodes the DFR protein.

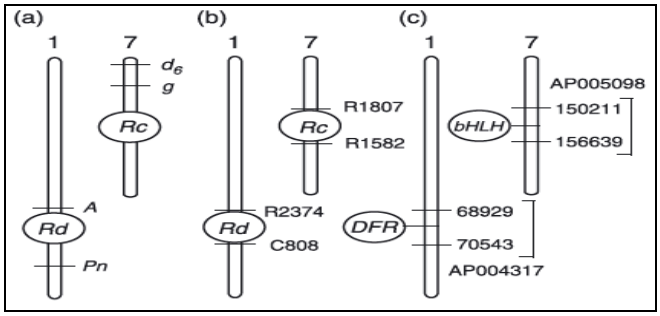
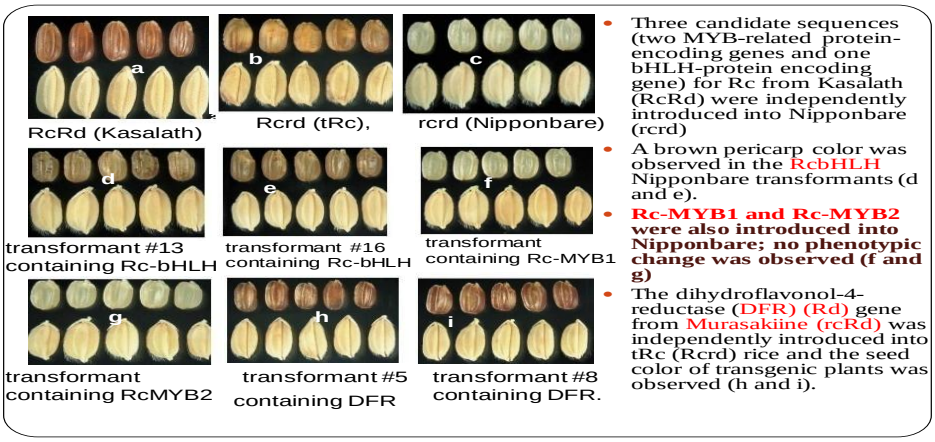


Fig 1: (a) Gene loci for *Rc* and *Rd* (b) Physical map of *Rc* and *Rd* and their adjacent DNA markers and (c) Physical location of the dihydroflavonol-4-reductase (DFR) and basic helix-loop-helix (bHLH) genes determined in this study.

Accumulation of proanthocyanidins was observed in the transformants by the introduction of the Rd gene into the rice Rcrd line. Protein blot analysis showed that the *DFR* gene was translated in seeds with alternative translation initiation. A search for the *Rc* gene, which encodes a transacting regulatory factor, were conducted using available DNA markers and the Rice Genome Automated Annotation System program. Three candidate genes were identified and cloned from a rice RcRd line and subsequently introduced into a rice rcrd line. Brown-colored seeds were obtained from transgenic plants by the introduction of a gene containing the basic helix–loop–helix (bHLH) motif, demonstrating that the *Rc* gene encodes a bHLH protein.

- *Rc* gene is responsible for the accumulation of pigments in the pericarp of brown- colored grains
- Red rice grains require the *Rd* gene to increase the content of the pigment



2. Identification of genes related to germination in aged maize seed by screening natural variability by Revilla *et al.* (2010) [13].

Ageing reduces vigour and viability in maize inbred lines due to non-heritable degenerative changes. Besides non heritable genetic changes due to chromosome aberrations and damage in the DNA sequence, heritable changes during maize conservation have been reported. Genetic variability among aged seeds of inbred lines could be used for association studies with seed germination.

Objective: To identify genes related to germination in aged seeds.

The sweet corn inbred line P39 and the field corn inbred line EP44 were used as plant material. Bulks of living and dead seeds after 20 and 22 years of storage were compared by using simple sequence repeats (SSRs) and, when the bulks differed for a marker, the individual grains were genotyped. Differences between dead and living seeds could be explained by residual variability, spontaneous mutation, or ageing. Variability was larger for chromosome 7 than for other chromosomes, and for distal than for proximal markers, suggesting some relationships between position in the genome and viability in aged seed. Polymorphic SSRs between living and dead seeds were found in six known genes, including pathogenesis-related protein 2, superoxide dismutase 4, catalase 3, opaque endosperm 2, and metallothionein1 that were related to germination, along with golden plant 2. In addition, five novel candidate genes have been identified; three of them could be involved in resistance to diseases, one in detoxification of electrophillic compounds, and another in transcription regulation. Therefore, genetic variability among aged seeds of inbreds was useful for preliminary association analysis to identify candidate genes.

The distribution of variability among chromosomes did not deviate from randomness except for chromosome 7, which showed a larger rate of variability than the other nine chromosomes.

Table 1: Deviations from the random distribution of polymorphic SSRs in the inbred lines P39 and EP44

Chromosome	Total no. of SSR markers	Variable SSR bands	
		Observed	Expected
Chromosome 1	27	6	3.7
Chromosome 2	23	3	3.2
Chromosome 3	16	2	2.2
Chromosome 4	28	3	3.8
Chromosome 5	22	1	3.0
Chromosome 6	25	2	3.4
Chromosome 7	17	6	2.3
Chromosome 8	23	2	3.2
Chromosome 9	18	1	2.5
Chromosome 10	27	5	3.7
	χ^2 Total = 16.9		
Position distal	23	6	3.6
Position proximal	22	1	3.4
	χ^2 Total = 3.4**		

χ^2 significant, at *P=0.05 and **P=0.01, respectively.

b. Trancriptomics

Transcriptomics is the study of the transcriptom- the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in specific cell-using high-throughput methods, such as microarray analysis. Comparison of transcriptomes allows the identification of genes that are differentially expressed in distinct cell populations, or in response to different treatments.

Central dogma of life

The central dogma of molecular biology is an explanation of the flow of genetic information within a biological system. It was first stated by Francis Crick in 1958 and re-stated in a *Nature* paper published in 1970 The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid. This has also been described as "DNA makes RNA makes protein. However, this simplification does not make it clear that the central dogma as stated by Crick does not preclude the reverse flow of information from RNA to DNA, but only the reverse flow from protein to RNA or

DNA. Crick had misapplied the term "dogma" and Crick's proposal had nothing to do with the linguist meaning of "dogma". He subsequently documented this error in his autobiography.

Technologies

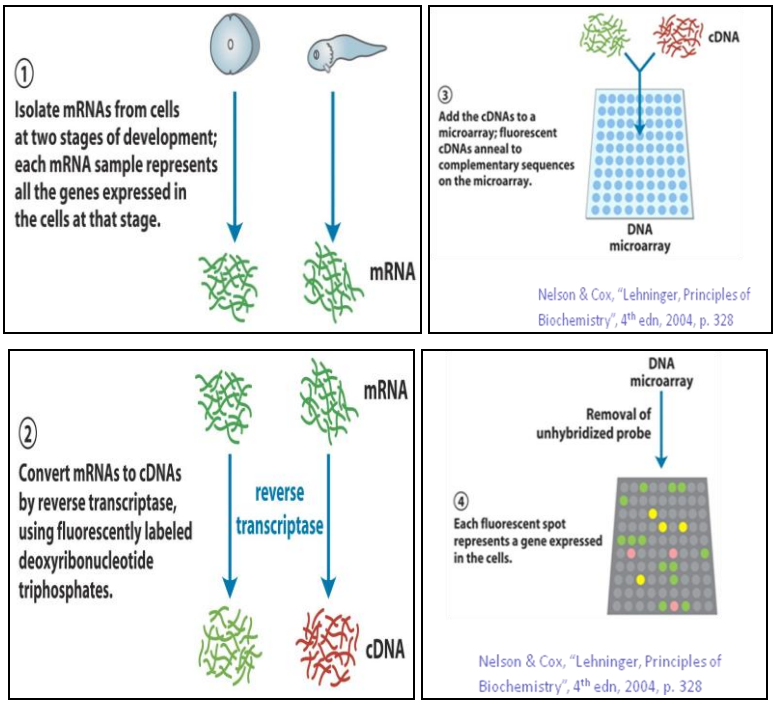
i. Hybridization-based approaches

- Fluorescently labelled cDNA with custom-made microarrays
- Commercial high-density oligo microarrays

ii. Sequence-based approaches

- SAGE-Serial analysis of gene expression
- CAGE- Cap analysis of gene expression
- RNA sequencing

Microarray



1. Identification and comparative analysis of drought associated microRNAs in two cowpea genotypes by Figueroa *et al.* (2011) [5]

Background: Cowpea (*Vigna unguiculata*) is an important crop in arid and semi-arid regions and is a good model for studying drought tolerance. MicroRNAs (miRNAs) are known to play critical roles in plant stress responses, but drought-associated miRNAs have not been identified in cowpea. In addition, it is not understood how miRNAs might contribute to different capacities of drought tolerance in different cowpea genotypes.

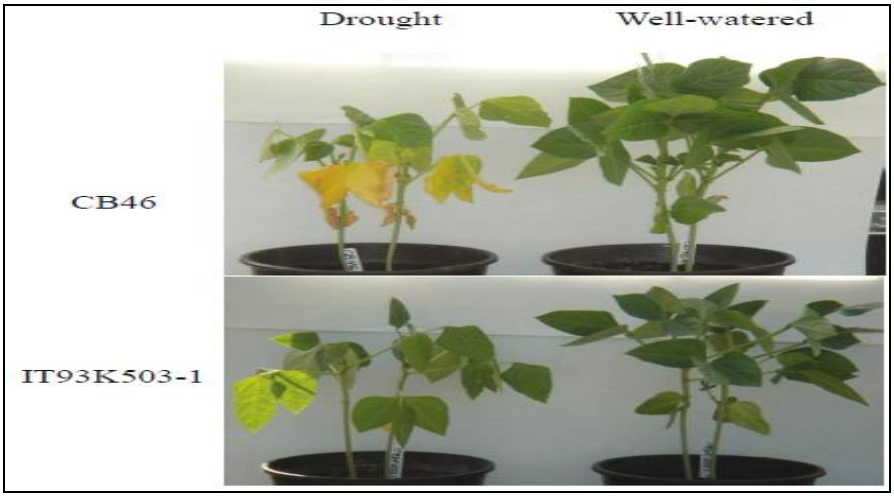


Fig: Different drought tolerance of two cowpea genotypes

Results: They generated deep sequencing small RNA reads from two cowpea genotypes (CB46, drought-sensitive, and IT93K503-1, drought-tolerant) that grew under well-watered and drought stress conditions. They mapped small RNA reads to cowpea genomic sequences and identified 157 miRNA genes that belong to 89 families. Among 44 drought-associated miRNAs, 30 were upregulated in drought condition and 14 were downregulated. Although miRNA expression was in general consistent in two genotypes, they found that nine miRNAs were predominantly or exclusively expressed in one of the two genotypes and that 11 miRNAs were drought-regulated in only one genotype, but not the other.

Table 1: MIRNAS that showed genotype-specific expressin

Family	Mature miRNA	Normalized Expression Level (TPTM)*				Putative Target
		IT93K503-Control	IT93K503-Drought	CB46-Control	CB46-Drought	
vun_cand058	UUAAGCAGAAUGAUCAAAUUG	942	1546	3	0	hydroxyproline-rich glycoprotein
vun_cand048	UGGUCUCUAAACUUUAGAAAUGAA	746	263	0	2	
vun_cand036	UCAGAGGAAACAACACUUGUAC	59	23	0	0	
vun_cand045	CGUGCUGAGAAAGUUGCUUCU	52	79	14	5	VTC2 (vitamin c defective 2)
vun_cand053	GUAAUUGAGUUAAAAGGACUAUAU	43	6	0	2	cellulose synthase/transferase
vun_cand052	CGAGAGCCACUCGCCUAAGCGA	34	55	0	0	
vun_cand055	CCACUGUAGUAGCUCUCGCUCA	30	40	0	0	
vun_cand054	AGCAAGUUGAGGAUGGAGCUU	9	48	231	252	CKA1 (casein kinase alpha 1)
vun_cand014	UUCGGGAGUGAGAGCCAGUGA	3	0	56	5	UBP18 (ubiquitin-specific protease 18)

*TMT transcripts per ten million

Conclusion: These results suggest that miRNAs may play important roles in drought tolerance in cowpea and may be a key factor in determining the level of drought tolerance in different cowpea genotypes.

2. Loss of unction of *OsDCL1* affectts microRNA accumulation and causes developmental defects in rice by Liu *et al.* (2005) [9]

Objectice: To study the function of rice DCL proteins (OsDCLs) in the biogenesis of miRNAs and siRNAs

- An RNA interference approach was applied to knock down two OsDCLs, OsDCL1 and OsDCL4, respectively

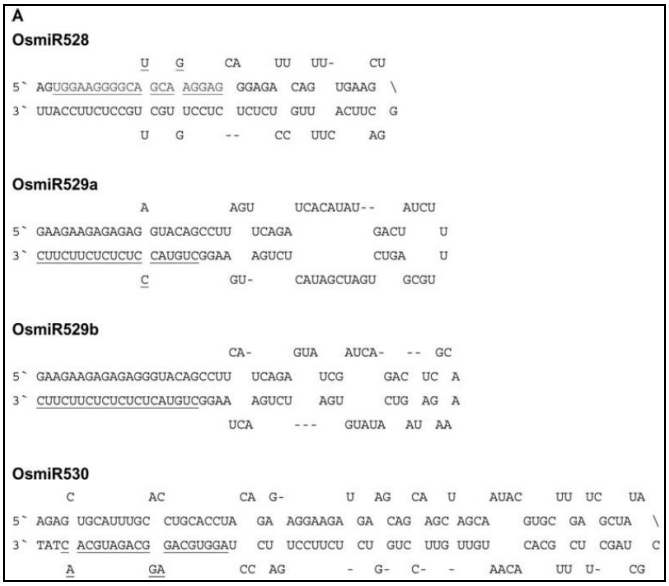
Materials and Methods

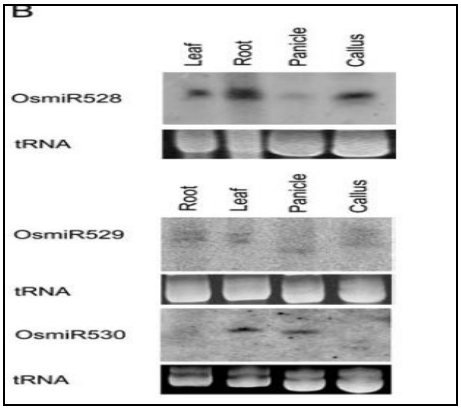
- Rice (Oryza sativa) japonica cv Nipponbare
- Small RNA Cloning, miRNA Prediction & Phylogenetic Analysis

Construction of RNAi Vector: pCam23ACT: OCS, a derivative of pCambia 2300, carrying the rice Actin1 promoter and the OCS terminator, was used for plant transformation

Results

1. Analysis of Small RNAs from Rice: Representatives of newly identified miRNAs from rice





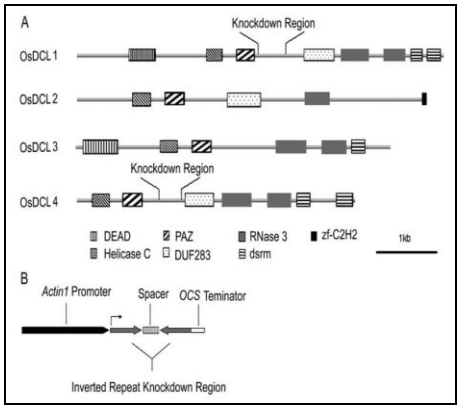
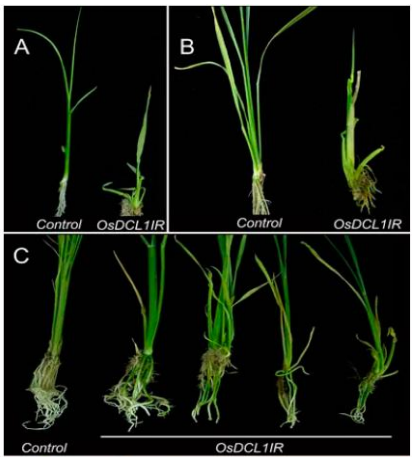


Fig: Expression patterns of novel rice miRNAs, A) Schematic representation of conserved motifs among four DCL proteins in rice and B) Diagram of OsDCL RNAi constructs



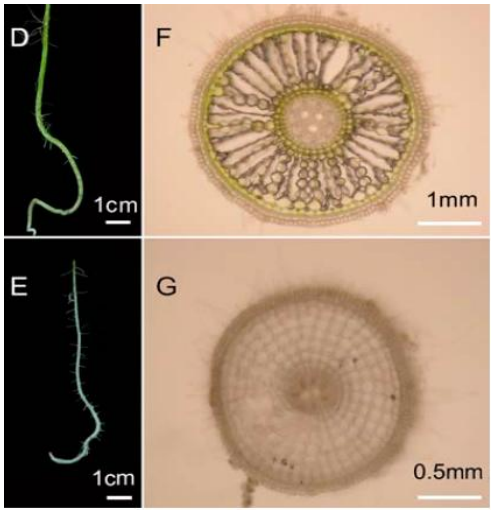


Fig: Wild type and strong loss of function of OsDCL1IR transformants

- Specificity of OsDCL1 for miRNA biogenesis is conserved between monocots and dicots.
- Functional study of OsDCLs demonstrated that OsDCL1 is required for miRNA processing and normal development.
- Loss of function of OsDCL1IR transformants greatly reducedmi RNA accumulation resulting in pleiotropic phenotypes.

c. Proteomics

Proteomics, the large-scale study of the total complement of proteins in a given sample, has been applied to all aspects of seed biology mainly using model species such as Arabidopsis or important agricultural crops such as corn and rice.

Proteins extracted from the sample have typically been separated and quantified by 2-dimensional polyacrylamide gel electrophoresis followed by liquid chromatography and mass spectrometry to identify the proteins in the gel spots. In this way, qualitative and quantitative changes in the proteome during seed development, desiccation tolerance, germination, dormancy release, vigor alteration and responses to environmental factors have all been studied.

Seed development: The seed develops from a single fertilized zygote into an embryo and endosperm in association with the surrounding maternal tissues. Most seeds contain large quantities of nutrient reserves, mainly carbohydrates, oils, and/or proteins, which are biosynthesized and deposited during seed development. These reserves are not only important for seed germination and seedling growth, but are also vital components of human and animal diets. Their production in crops is the basis of agriculture. Before reaching maturity, the seed develops other important properties, including desiccation tolerance, germination/dormancy and vigor (Bewley *et al.*, 2013) [2].

Seed desiccation tolerance: Considered only in terms of tolerance of, or sensitivity to, desiccation, seeds can be categorized as orthodox or recalcitrant. The orthodox seed acquires desiccation tolerance during seed development approximately halfway through development. This trait ensures that the seeds passes unharmed through maturation drying and retains viability in the dry state for long periods of time (up to hundreds of years in some case) under natural or artificial conditions.

Seed germination: Seed germination is the most critical phase in the seed plant life cycle. It determines when the plant enters natural or agricultural ecosystems. Cultivation of most crop species is dependent on seed germination. Seeds of most species acquire the ability to germinate during development. This is important for crop production, because it ensures that the untreated seed quickly germinates after sowing. However, in a few species, such as maize, wheat and rice, it can result in precocious germination, which typically occurs when developing seeds with a low degree of dormancy experience rainfall or humid conditions. Precocious germination can decrease the grain quality and cause great economic losses. In some species, seeds are dormant at the end of development.

Seed dormancy: is defined as the failure of an intact viable seed to complete germination under favorable conditions (Bewley, 1997) [2]. It is an adaptive strategy for seed to survive under adverse natural conditions, but it also creates an obstacle for agricultural production, where rapid germination and growth are required.

Seed proteomics: The great biological and economic importance of seeds has led to a vast number of studies of all the above aspects of seed biology. One type of study is proteomics, the study of all the expressed proteins. Since proteins are responsible for most metabolic processes in the seed, in addition to being important structural components in the cytoskeleton, membranes, the cell wall, etc., it makes excellent sense to describe the proteome of a seed, a seed tissue, a specific cell type or a subcellular compartment. However, proteomics are also a powerful tool for detecting changes in the protein composition in response to developmental or environmental stimuli, so-called differential proteomics.

Proteomic methods

The seed proteome can be analyzed like any other proteome using the standard general procedure of protein extraction, separation and identification.

Gel-based methods for separation particularly two dimensional polyacrylamide gel electrophoresis (2D-PAGE) have dominated and will probably continue to dominate because, In addition to being reasonably quantitative, they provide a lot of information about the proteins not provided by the gel-free shotgun methods, such as changes in protein size, pI, and posttranslational modifications (PTMs).

Mass spectrometer-based methods for protein identification have dominated completely in recent years also in seed proteomics, but the methods have been extensively reviewed recently (*e.g.* Pan *et al.*, 2009; Walther and Mann, 2010; Bantscheff *et al.*, 2012) [11, 15, 1] and since they are essentially independent of species (except for the question of access to a full genome sequence), they will not be reviewed here. 1.2. Pitfalls in the use of 2D-PAGE and other quantitative proteomic methods although 2D-PAGE is reasonably quantitative, a note of caution is in order concerning the way it is routinely used.

Protein synthesis

- Cell division and enlargement occurring during seed development require proteins involved in processes such as DNA replication, transcription, cytoskeleton and cell wall formation.
- In many species, most of the proteins involved in protein synthesis including ribosomal proteins, translation factors

and tRNA synthases, in protein folding and stability, such as heat shock proteins

Reserve accumulation

- Most mature seeds contain two or more reserve compounds including carbohydrates, oils and proteins, and to a large extent they are synthesized during seed development, especially during the stage of reserve deposition.
- Sucrose and amino acids, imported from the parent plant, are the major carbon and nitrogen sources for the synthesis of reserve compounds.

Energy production: The central metabolism (glycolysis and tricarboxylic acid (TCA)

- The central metabolism (glycolysis and tricarboxylic acid (TCA) cycle) provides most of the energy for processes in the seed.
- In developing seeds, most of the glycolytic enzymes in the cytosol and a few in plastids.
- In castor seeds nearly all the glycolytic enzymes in both cytosol and plastids, with the exception of plastidic phosphoglyceromutase, were identified and showed change in abundance during seed development.
- Proteomic data gave what appeared to be a relatively clear picture of the involvement of glycolysis during seed development and identified glycolytic enzymes vary greatly not only in number, but also in accumulation pattern, among different species

In developing seeds, most of the glycolytic enzymes in the cytosol and a few in plastids have been identified in many species by proteomics. In castor seeds nearly all the glycolytic enzymes in both cytosol and plastids, with the exception of plastidic phosphoglyceromutase, were identified and shown to change in abundance during seed development. These data gave what appeared to be a relatively clear picture of the involvement of glycolysis during seed development. However, when similar studies were performed in other species, it became clear that the identified glycolytic enzymes vary greatly not only in number, but also in accumulation pattern, among different species. For example, using the same 2-D PAGE proteomic approach, a total of 63 protein spots involved in glycolysis were identified in developing castor seeds, while this number was only 19 in Arabidopsis seeds (Hajduch *et al*., 2010) [7]. In castor seeds, the identified glycolytic proteins in both cytosol and plastids accumulated most abundantly at the stage of histodifferentiation and decreased gradually during reserve deposition, while in Arabidopsis, various trends of change were observed. A variation in the number and accumulation pattern was also observed for the enzymes of the TCA cycle.

The above results imply that the participation of glycolysis and the TCA cycle in seed development is not entirely straightforward and possibly regulated by different mechanisms in different plant species. In addition to providing energy in the form of ATP glycolysis and the TCA cycle also provide many intermediates for the biosynthesis of storage reserves, secondary metabolites, nucleotides, etc., where the need differs among different plant species. Therefore, the various accumulation patterns of proteins involved in glycolysis and the TCA cycle may reflect different requirement for glycolytic and/or TCA cycle intermediates in biosynthesis.

Desiccation tolerance

- Desiccation tolerance refers to the ability of an organism to endure loss of a almost all of its cellular water without irreversible damage.
- In higher plant, only orthodox seeds are desiccation tolerant. This property gives the orthodox seeds the ability to survive under extreme environmental conditions.

Proteomic studies of seed desiccation tolerance

Accumulation of LEA proteins

- Late embryogenesis abundant (LEA) proteins are well characterized as protective molecules against desiccation stress (Cuming, 1999) [4].
- They are thought to act by replacing water, sequestering ions, removing ROS and/or stabilizing protein and membrane structure.
- Six LEA proteins from four gene groups, including Em6, MP2, an isoform of PM18, six isoforms of SBP65, PM25, and one isoform of DHN3 were identified to be associated with desiccation tolerance in *M. truncatula* seeds.
- This further validates a correlation between the absence of LEA protein accumulation and seed desiccation sensitivity. This study also revealed that most of the desiccation tolerance-related LEA proteins were positively regulated by ABI3.

Removal of Reactive Oxygen Species

Dehydration will disrupt the metabolism of seeds and lead to production of ROS, such as H_2O_2, O_2-, singlet oxygen and the hydroxyl radical. At lower concentration, ROS can act as a messenger to regulate biological process, while they can damage cellular components, like lipids, proteins and DNA at higher concentration. It is also possible that breakdown products, *e.g.* oxidized peptides deriving from oxidized proteins, can act as signals (Moller and Sweet love, 2010) [10]. Thus, ROS and the various oxidation products must be strictly controlled in seeds during dehydration.

Stabilization of structure

In the pea seed proteome, the amounts of two proteins related to structural stabilization, the TCP-1/cpn60 chaperonin family protein and tubulin a-1 chain decreased during seed germination. Pea seeds imbibed in $CaCl_2$ and MV were more and less tolerant to dehydration, respectively, compared to seeds imbibed in distilled water.

Proteomic changes associated with seed dormancy

Dormancy is the temporary failure of a seed to complete germination under favorable conditions. This means that they cannot complete phase II. "According to the hormone balance theory, the relative actions of abscisic acid (ABA) (inhibitory) and gibberellic acid (GA) (promotive) are the primary determinants of seed dormancy and germination". However, a proteomic study in Arabidopsis showed that the proteomic profile of dormant seeds was quite different from that of non-dormant seeds treated with exogenous ABA to make them dormant. This indicates that the mechanism of dormancy induction also differed.

About 35 proteins belonging to the groups of storage proteins and stress response and detoxification were more abundant in the dormant seeds, while proteins belonging to the groups of energy (20), amino acid metabolism, folding and stability, proteolysis and mRNA metabolism and protein synthesis

(about 10 each) were more abundant in the non-dormant seeds.

Changes in post-translational modifications during germination

PTMs in proteins can be regulatory, *e.g.* phosphorylation, they can be part of protein degradation pathways, *e.g.* ubiquitination, or they can perhaps be both as proposed for carbonylation. Protein carbonylation occurs as a result of metal-catalyzed oxidation involving ROS and it is probably the most common irreversible protein oxidation PTM. Han *et al*. identified more than 800 phosphoproteins in rice seeds, out of which 149 changed in amount during germination.

Seed vigor

Seeds with different vigor resulting from aging and priming, where aging decreases and priming increases seed vigor, have been studied by proteomics. These studies have led to the identification of many proteins and metabolic processes potentially important for seed vigor.

1. Proteome analysis of grain filling and seed maturation in barley by Finnie *et al.*, (2016)

Field-grown seeds were collected at weekly intervals over a period of 5 weeks during the grain filling and maturation stage. According to the Zadoks scale of cereal grain development, the five samples of barley cv Barke corresponded to the following stages, respectively: 80 (start of dough development), 82, 85 (soft dough, onset of drying), 86, and 87 (hard dough). Seeds from three other barley cultivars with varying properties were likewise collected for comparison because they would be expected to differ in the protein expression patterns visible by two-dimensional gel electrophoresis.

- Protein spots were divided into six categories according to the timing of appearance or disappearance during the 5-week period of comparison.
- Changes in individual protein spots over the seed development period enabled classification according to expression patterns

  Group-O, Present at all stages
  Group-I, Early, decreasing during development
  Group-II, Increasing during development
  Group-III, Transient
  Group-IV, Mid, coincident with dessication
  Group-V, Late
  Group-VI, Other patterns

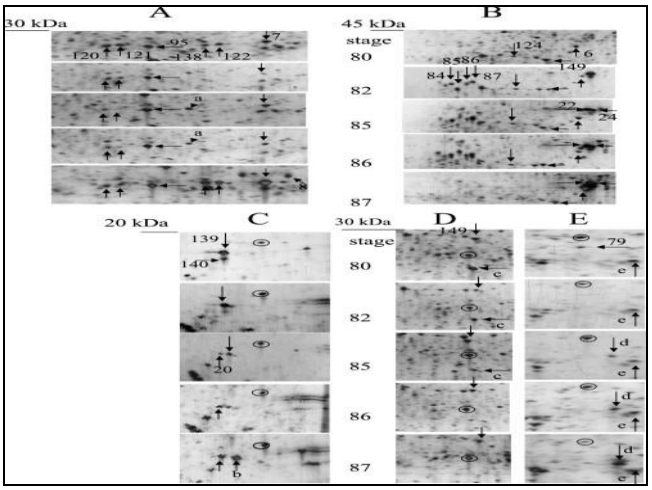
A. This region shows a group of spots that is abundant at the early and late stages of dough formation but less abundant at the middle stage (upwards arrows). Another spot remains constant throughout the process (horizontal arrow) and yet another spot is present at all stages of development, but most abundant at the final stage (downwards arrow). Spots 7, 9, 95, 120 through 122, and 138 have been identified as triose phosphate isomerase. An as yet unidentified spot that has expression pattern III is marked (a).

B. Group of spots (spots 84–87) accumulates transiently at the middle stages of development. These have been identified as C-terminal fragments of protein disulfide isomerase (PDI). Another group of spots, including spots 22 and 24, increases gradually both in intensity and number throughout development and are identified as protein Z4 (a Ser protease inhibitor [serpin]). Below the serpin spots, glyoxalase I (spot 6) is present throughout and is particularly abundant at the latest stage. A fragment of this protein is also present (spot 149).

C. Spot 140 is probably a degradation product of spot 139. These spots decrease gradually in intensity during development (group I). Both have been identified as the small subunit of ribulose bisphosphate carboxylase. Spot 20 (group IV, desiccation related) is a glyoxalase I-related protein (see text). Another, unidentified protein (spot b) has the same pattern of appearance.

D. Spot 149 (glyoxalase I fragment) appears to alter position during development. An as yet unidentified spot with expression pattern I (c) is marked.

E. Spot 79, also with expression pattern I, has been identified as cytosolic ascorbate peroxidase (APX). Spots with expression pattern IV (d) and O (e) are marked. Spot d has been identified as a 1cys peroxiredoxin, whereas spot e is not yet identified.



Variations in protein spots during seed development

b. Metabolomics

Metabolites: Small molecules that participate in metabolic reactions and required for growth and normal function of a cell.

Metabolome: The complete set of metabolites in an organism.

What is Metabolomics?

Metabolomics is an advanced, specialised form of analytical biochemistry. The technology focuses on the detection of small molecules which often play key roles in *e.g.* food quality, disease resistance, antioxidant activity etc. Important small metabolites include *e.g.* organic acids, amino acids, sugars, volatile metabolites and a myriad of secondary metabolites such as alkaloids, phenolic components and also pigments such as carotenoids and anthocyanins.

Chromatography (Liquid or Gas) are usually used to separate complex plant extracts into the individual components; Mass Spectrometry (MS) is then used to detect and if possible, quantify the metabolites present. Alternatively, Nuclear Magnetic Resonance (NMR) may be used. Metabolomics is primarily distinguished from more established technologies by the high throughput nature of the approach which also generates a complex dataset per analysis. As a consequence, the technology relies heavily on recent advances made in

bioinformatics and Information Technology in order to process, store and mine the complex data matrices for biological information. Metabolomics is usually used either for 'fingerprinting' samples to perform comparative analyses to detect differences of for 'profiling' where individual differential metabolites, perhaps linked to key quality traits, are identified for further analysis.

Breeding Healthier Foods

In most diets, plant products provide the main part of human food intake, and the link between food and health is becoming increasingly clear. Metabolomics can play an important role in improving the healthy value of plants. With metabolomics, a comprehensive set of biomarkers can be generated, which can be used to understand and monitor the interaction between food intake/uptake and human health. These data can provide essential indicators as to how food should be produced and offered to meet modern health needs. Metabolomics is predicted to become a cornerstone in this field. Metabolomics shall be employed to direct breeding strategies to enhance specific desired balances of food components in fresh food which have been identified as being more optimal. In addition, the technology shall prove pivotal to the further optimisation of food processing methods to help generate or preserve the desired nutritional balance in longer shelf-life strategies. However, before all this is possible in the most optimal way, the technology requires further development and fine-tuning.

Metabolomics and shelf life

Having a better understanding of which processes are involved in product deterioration, and how these might be counter-acted, increases the supplier's ability to predict the maximum shelf life of a particular batch of product more reliably and to *e.g.* compensate for environmentally or source – related batch differences. In the META-PHOR project for example, it could be shown that the transition between material before/after the quality limit is very short and the switch to a lower quality product (with developing off-flavours etc.) occurs in a rapid time-frame. The quest is on to link this transition to key predictors (biomarkers) present in the initial starting material as it arrives at the processing factory. A series of such biomarkers, identified through metabolomics, can then be used for a batch-wise determination of shelf-life.

Potato Metabolomics

In particular, the development of a better potato, via the introgression of different metabolite contents and diversities, with a view to developing new varieties has been addressed by different groups working with wild species collections. For example, the Commonwealth Potato Collection – 1,500 accessions of about 80 wild and cultivated potato species – was sub-sampled and analysed by Gas Chormatography-Mass Spectrometry (GC-MS) focussed metabolomics. These analyses clearly showed that specific taxonomic groups segregated on the basis of their biochemical composition (*e.g.* amino acids, sugars). Others have utilized alternative metabolomics technologies, to assess compositional differences in existing potato cultivars. The levels of the amino acids isoleucine, tyrosine and phenylalanine were found to be higher in certain cultivars- an important finding as these amino acids are associated with flavour/aroma, post cooking blackening and bruising. In addition, metabolomics on potato is also revealing chemical links to quality issues not previously recognised, thus emphasising the value of this broad approach.

Tomato Metabolomics

Tomato is perhaps the most widely studied crop using metabolomics technologies. Investigations into varietal differences, cultivation influences, genetic modification have all greatly advanced our knowledge of the biochemical composition of the tomato fruit and how this is altered by environment and genetics. Key components playing determinant roles in taste characteristics- both positive and negative – have been identified and can be used as quality predictors. The influence of seasonality and growth conditions on fruit quality has revealed extensive, previously unknown changes and help us explain better why *e.g.* batch/source differences are so frequently observed even when using the same hybrid variety.

1. Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars by Lin *et al*. (2014) [8]

Soybean [*Glycine max* (L.) Merr.] is one of the world's major crops, and soybean seeds are a rich and important resource for proteins and oils. In this study, they investigated the seed metabolomes of 29 common soybean cultivars through combined gas chromatography-mass spectrometry and ultra-performance liquid chromatography-tandem mass spectrometry. One hundred sixty-nine named metabolites were identified and subsequently used to construct a metabolic network of mature soybean seed. Among the 169 detected metabolites, 104 were found to be significantly variable in their levels across tested cultivars. Metabolite markers that could be used to distinguish genetically related soybean cultivars were also identified, and metabolite–metabolite correlation analysis revealed some significant associations within the same or among different metabolite groups. Findings from this work may potentially provide the basis for further studies on both soybean seed metabolism and metabolic engineering to improve soybean seed quality and yield.

The heat map revealed remarkable diversities in metabolite abundance across all cultivars. The 169 metabolites could be clustered into 10 classes. Each class of metabolite was enriched in different cultivars. For example, clas-1 contained 18 metabolites that were highly represented in cultivars 1138-2 and HCD. Class-2 contained 17 metabolites that were only abundant in cultivar DBH, and 13 of the 17 metabolites in this class were related to lipid or secondary metabolism. The seven metabolites in group 9 were highly enriched in cultivar PSHD. Furthermore, class 3 (17 metabolites) and class 4 (12 metabolites) were both high in cultivar S-1. In contrast to the cultivars with enriched metabolite classes, several soybean cultivars, such as 58-161, Williams, and WDZ, appeared to have lower abundance of most of the detected metabolites when compared to the other cultivars. In the hybrid cultivar Su-1, many metabolites displayed higher abundance than in the parents (Nannong 493-1 and 58-161). Interestingly, 1138-2 and 58-161, two cultivars obtained by natural selection, displayed different metabolite changes when compared to their parents: many metabolites in 1138-2 were increased compared to its parent SDH, whereas many metabolites in 58-161 were decreased compared to its parent DBH. Finally, some individual metabolites were enriched only in specific cultivars

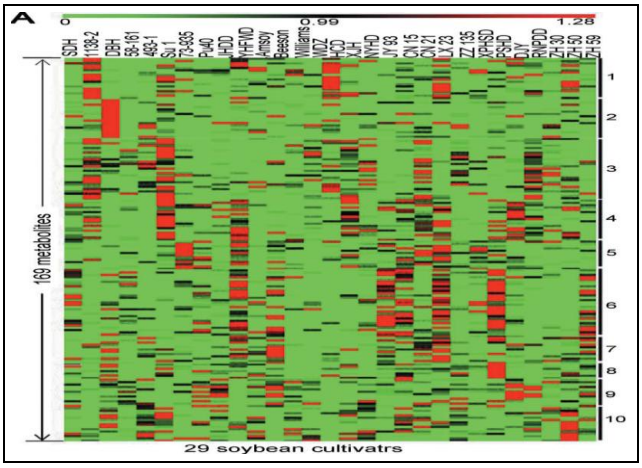


Fig: Clustering analysis of the metabolic data of all soybean cultivars

Conclusion

The twenty-first century omics technologies play an essential role for improved seed development. Omics in seed development have provided biologists with comprehensive data concerning global gene function and identification, dynamic changes between gene expression, protein identification, function and interaction of entire protein networks.

References

1. Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal. Bioanal. Chem. 2012; 404:939-965.
2. Bewley JD. Seed germination and dormancy. Plant Cell, 1997; 9:1055-1066.
3. Bewley JD, Bradford KJ, Hilborst HWM, Nonogaki H. Seeds –Physiology of development, germination and dormancy, third ed. Springer, New York, 2013.
4. Cuming AC. LEA proteins. In: Shewry, P.R., Casey, R. (Eds.), Seed Proteins. Kluwer Academic Press, Dordrecht, The Netherlands, 1999, 753-780.
5. Figueroa BEB, Gao L, Diop NN, Wu Z, Ehlers JD, Roberts PA *et al*. Identification and comparative analysis of drought-associated microRNAs in two cowpea Genotypes. BMC Plant Biology. 2011; 11:127-130.
6. Furukawa T, Maekawa M, Oki T, Suda I, Iida S, Shimada H *et al*. The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. The Plant J., 2006; 49:91-102.
7. Hajduch M, Hearne LB, Miernyk JA, Casteel JE, Joshi T, Agrawal GK *et al*. Systems analysis of feed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression. Plant Physiol. 2010; 152:2078-2087.
8. Lin H, Jun R, Jianxin S, Chaoyang H, Fang C, Zoe A *et al*. Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. J. Integrative Pl. Bio. 2014; 56(9):826-836.
9. Liu B, Ping Chuan Li, Xin Li, Chun Yan Liu, Shou Yun Cao, Cheng Cai Chu *et al*. Loss of function of OsDCL1 affects MicroRNA accumulation and causes developmental defects in rice. Plant Physiol. 2005; 139:296-305.
10. Moller IM, Sweetlove LJ. ROS signalling e specificity is required. Trends Plant Sci. 2010; 15:370-374.
11. Pan S, Aebersold R, Chen R, Rush J, Goodlett DR, McIntosh MW *et al*. Mass spectrometry based targeted protein quantification: methods and applications. J. Proteome Res. 2009; 8:787-797.
12. Radchuk V, Borisjuk L. *Jekyll* encodes a novel protein involved in the sexual reproduction of barley. *Plant Cell*, 2006; 18:1652-1666.
13. Revilla P, Butron AV, Rodriguez M, Malvar RA, Orda A. Identification of genes related to germination in aged maize seed by screening natural variability. J. Exp. Bot. 2010; 14:4151-4157.
14. Saito K, Matsuda F. Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol*, 2010; 61:463-489.
15. Walther TC, Mann M. Mass spectrometry-based proteomics in cell biology. J. Cell. Biol. 2010; 190:491-500.
16. Weber H, Borisjuk L, Wobus U. Molecular physiology of legume seed development. *Annu*. Rev. Plant Biol. 2005; 56:253-279.
17. Xu H, Wei Y, Zhu Y, Lian L, Xie H, Cai Q *et al*. Antisense suppression of *LOX3* gene expression in rice endosperm enhances seed longevity. Plant Bio. J. 2015; 13:526-539.