Proteomic analysis of endosperm and peripheral layers during kernel development of wheat (Triticum aestivum L.) and a preliminary approach of data integration with transcriptome

Ayesha Tasleem Tasleem-Tahir

UNIVERSITE BLAISE PASCAL **UNIVERSITE D'AUVERGNE**

N° D.U. 2252 ANNEE 2012

ECOLE DOCTORALE DES SCIENCES DE LA VIE, SANTE, AGRONOMIE, ENVIRONNEMENT
N° d'ordre 585

Thèse

Présentée à l'Université Blaise Pascal
Pour l'obtention du grade de

DOCTEUR D'UNIVERSITE
Spécialité: Physiologie et génétique moléculaire

Soutenue le 4 Juillet 2012

Ayesha TASLEEM-TAHIR

Analyse du protéome de l'albumen et des couches périphériques du grain de blé (*Triticum aestivum* L.) en développement: vers une intégration des données avec le transcriptome

Proteomic analysis of endosperm and peripheral layers during kernel development of wheat (Triticum aestivum L.) and a preliminary approach of data integration with transcriptome

---

Président :	Saïd MOUZEYAR	Professeur, UBP, Clermont Ferrand, France
Rapporteurs :	Birte SVENSSON	Professeur, DTU, Lyngby, Denmark
	Dominique JOB	Directeur de Recherche, CNRS/BAYER, Lyon, France
Examinateur :	Didier MARION	Directeur de Recherche, INRA, Nantes, France
Directeur de thèse :	Gérard BRANLARD	Directeur de Recherche, INRA, Clermont Ferrand, France

Unité Mixte de Recherche 1095 INRA-Université Blaise Pascal
Génétique, Diversité et Ecophysiologie des Céréales
63039 Clermont Ferrand

ABSTRACT / RESUME

Proteomic analysis of endosperm and peripheral layers during kernel development of wheat (*Triticum aestivum* L.) and a preliminary approach of data integration with transcriptome

Wheat is the second most produced cereal in the world, important for food, feed and many industrial uses. Understanding of the mechanisms involved in grain development is fundamental for developing high quality wheat. In particular, detailed knowledge of the wheat grain physiology and molecular mechanisms involved in its development would help in breeding not only of wheat but also many other cereals. A proteomic approach has been used in this context but, up to now, there had been no work on developing tissues at very short temporal distances.

This thesis presents, firstly, a proteomic study to characterize protein expression changes in peripheral layers and in starchy endosperm of wheat, during kernel development. We used grains of *Triticum aestivum* cv Récital, cultivated at INRA, Clermont-Ferrand. Grains were harvested at each 50°Cd from fertilization to maturity at fifteen stages for peripheral layers and at twenty-one stages for starchy endosperm. After grain dissection, protein extraction and 2DE- MALDI-TOF MS and data mining, we identified 207 differentially expressed proteins at fifteen stages (0°Cd-700°Cd) of peripheral layers during kernel development. These proteins were then classed in sixteen different functional classes. HCA revealed five different expression profiles during development. Similarly after obtaining starchy endosperm from dissected grains, we performed protein extraction specific to metabolic proteins. After 2DE, 487 proteins were identified from fertilization to grain maturity (0°Cd-1006°Cd), using LC-MS and data mining. Proteins were grouped in nine different expression profiles and were classed in seventeen biochemical functions.

We have constructed proteome maps of these two important grain tissues during kernel development. Further, the comparison of peripheral layers and starchy endosperm proteomic data was made, with an objective to understand whether the changes in different biochemical processes differ between these tissues.
Finally, we performed an integration of our proteomic data (focusing our approach on proteins involved in carbohydrate metabolism) with that of transcriptomics. Only 32% of proteome/transcriptome expression profiles showed a significant correlation during development (from 152°Cd-700°Cd). Comparison of enzyme expression profiles with those of proteome and transcriptome would help to distinguish the processes regulated at transcriptome level and those controlled at the proteome level. This comprehensive grain development data could further help in construction of a Récital databank, which may be used as reference for studies of diseased and stressed grain tissues during development.

Keywords: *Triticum aestivum*, grain development, peripheral layers, starchy endosperm, metabolic proteins, expression profiles, proteome/transcriptome

Analyse du protéome de l'albumen et des couches périphérique du grain de blé (*Triticum aestivum* L.) en développement: vers une intégration des données avec le transcriptome

Le blé est la seconde céréale la plus produite dans le monde. Il constitue une importante source de denrées alimentaires et de beaucoup d'autres usages industriels. La compréhension des mécanismes impliqués dans le développement du grain de blé est fondamentale pour développer des blés à valeur ajoutée. La physiologie du grain de blé et les mécanismes moléculaires impliqués dans son développement nécessitent d'être mieux connus et ces connaissances pourront être très utiles pour l'amélioration du blé mais aussi des autres céréales. L'approche protéomique a été aussi utilisée dans ce contexte mais aucun travail n'avait jusqu'ici été réalisé sur la totalité des phases de développement des tissus et sur des intervalles de temps très courts. La caractérisation des changements d'expressions protéiques dans les couches périphériques du grain et de l'albumen est présentée dans cette étude.

Nous avons utilisé les grains de *Triticum aestivum* de la variété Récital, cultivés à l'INRA de Clermont-Ferrand. Les grains ont été prélevés tous les 50°C jour (°Cj) depuis la fécondation jusqu'à la maturité sur 15 stades de développement pour les couches périphériques et sur 21 stades pour l'albumen amylacé. Pour chaque échantillon, les couches périphériques des grains ont été disséquées et les protéines totales extraites. L'analyse des protéines en électrophorèse bidimensionnelle puis par spectrométrie de masse MALDI-TOF a permis d'identifier via l'interrogation des bases de données, 207 protéines différentiellement exprimées sur 15 stades de développement (0°Cj-700°Cj). Ces protéines ont ensuite été classées en 16 classes fonctionnelles. L'analyse en cluster a révélé 5 profils d'expression au cours du temps. Parallèlement, l'albumen amylacé a été isolé des grains et les protéines métaboliques de ce tissu extraites. Après électrophorèse bidimensionnelle des protéines, 487 protéines variant significativement dans l'albumen sur l'ensemble des stades de développement (0°Cj-1006°Cj) ont été identifiées par utilisation de la LC-MS. Les protéines ont été réparties sur neuf profils d'expression et 17 fonctions biochimiques. Le protéome des couches periphériques a ensuite été comparé au protéome de l'albumen dans le but de comprendre si l'évolution des processus biochimiques diffère dans chacun de ces tissus.

Au final, nous avons optimisé l'intégration des données protéomiques avec celles du transcriptome (en se focalisant sur les protéines du métabolisme carboné). Seulement 32% des profils d'expression protéome/transcriptome montrent une corrélation significative au cours du développement (152°Cj-700°Cj). Les profils d'expression des enzymes ont été comparés sur les deux niveaux. Ils devraient permettre de distinguer les processus régulés au niveau du transcriptome de ceux régulés au niveau du protéome. L'ensemble de ces données pourra être compilé dans une base de données propre de la variété Récital et utilisé comme référence dans l'étude des maladies et des stress abiotiques des tissus du grain de blé en développement.

Mots-clés : *Triticum aestivum*, grain en développement, couches périphériques, albumen amylacé, protéines métaboliques, profils d'expression, lien des données protéome/transcriptome.

REMERCIEMENTS

ACKNOWLEDGMENTS

This dissertation would not have been possible without the guidance and the help of several persons who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

I take this opportunity to pay my humble regard to Dr. Gilles CHARMET (former director of the UMR, Génétique, Diversité et Ecophysiologie des Céréales) for introducing me to his elite group of scientists. I would like to record my gratitude also to Dr. Thierry LANGIN for providing me his support in various ways since he became director.

I offer my sincerest gratitude to my thesis director Dr. Gérard BRANLARD, who has supported me thoughout my thesis with his patience and knowledge whilst allowing me room to work in my own way. His dynamic guidance and innovative ideas gave my work an appealing turn. One simply could not wish for a better or friendlier supervisor and I am indebted to him and cannot repay him any humanly way possible.

I would like to thank Prof. Birte SVENSSON, Dr. Dominique JOB, Didier MARION and Saïd MOUZEYAR for accepting to examine my thesis.
I also acknowledge my thesis committee, for their guidance and fruitful comments regarding my work.

I am also grateful to Pierre MARTRE, the head of our team, for his constructive comments and ideas. I am indebted to my team colleagues particularly Isabelle NADAUD, Emanuelle BANCEL and Marielle MERLINO for their fruitful discussions and suggestions.
I would like to convey my gratitude to all BIG team members who supported me in every step and left good memories for me. In my daily work I have been blessed with a friendly group of fellow students, Anne, Jonathan, Helene, Huiling, Thibault. Especially I would like to thank Clement DEBITON for his guidance and MACROS, which helped a lot during my work.
I also thank Christophe CHAMBON and Didier VIALA for Mass Spectrometry analysis.

I want to convey my gratitude to Frederic CHOULET, for his cooperation, gentle behavior and assistance in my first steps in bioinformatics. I would like to thank Veronique SAYS, both for her assistance and for making it possible for me to be at INRA early in the morning.

I owe my deepest gratitude to Sunflower Team, Felicity VEAR and Sylvie ROCHE, who provided me space not only in their team but also for moral support and at coffee time with their baking competence.
Valérie, Karine and Patricia for their help in official documentation during my stay at INRA.

I am obliged to Higher Education Commission, Pakistan, who has given me opportunity for higher studies in France and the team of SFERE for helping me out in every way possible.

Finally, I would like to thank everybody who helped in the successful completion of thesis.

.

TABLE OF CONTENTS

FIGURES AND TABLES

ABBREVIATIONS

AI	ATP Interconversion
AL	Aleurone Layer
Alg	Albumins-globulins
AM	Amino acid Metabolism
ANOVA	Analysis of variance
BSM	Biosynthesis of Secondary Metabolism
°CD	Degrees days after anthesis / pollination
CD	Cell Division
CK	Cytoskeleton (protein category)
CM	Carbohydrate Metabolism
Cysk	Cytoskeleton (sub-cellular localization)
Cyto	Cytosol
CW	Cell Wall related
DAA	Days After Anthesis
DAF	Days After Flowering
DAP	Days After Pollination
EM	Energy Metabolism
ER	Endoplasmic Reticulum
ESM	Endosperm
EST	Expressed Sequence Tag
Extr	Extracellular region
FM	Folic acid Metabolism
HCA	Hierarchical Clustering Analysis
HPLC	High Performance Liquid Chromatography
IEF	Isoelectrofocalisation
IPG	Immobilized pH Gradient
LC-MSMS	Liquid Chromatography coupled to tandem Mass Spectrometry
LM	Lipid Metabolism
M	Mitochondria

MALDI-TOF	Matrix Assisted Laser Desorption/Ionisation – Time of Flight
MT	Membrane Transport
mRNA	Messenger Ribonucleic Acid
NM	Nucleotide Metabolism
nucl	nuclear
P	Plastid
Perox	Peroxisome
PF	Protein Folding
pI	Isoelectric point
PL	Peripheral Layers
PM	Plasma Membrane
PR	Protein
PS	Protein Synthesis
PTO	Protein Turn Over
SD	Stress/Defense
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SP	Storage Protein
ST	Signal Transduction
tBlastn	Protein query - Translated database
TCA	Tricarboxylic Acid Cycle
TR	Transcript
Trs	Transcription/translation
UP	Unknown Proteins
2DE	2 Dimensional Electrophoresis

SUMMARY

Wheat is one of the major source of energy and protein for human beings as it is the second most important cereal crop grown in the world. During the last 30 years, yield had increased almost threefold (30 to 80 qx/ha). However, genetic constructions and agro-environmental conditions have led to a negative correlation between yield and protein content (Oury et al. 2003). To remedy this problem, significant progress has already been made on the genetics of grain storage proteins and their role in wheat quality (Branlard G and Marion D 2011). The aims of present research are not only to produce new varieties with higher protein content but also to control variability in the composition of components involved in the nutritional value of wheat.

The project of the UMR 1095 (Unité Mixte de Recherche) at INRA Clermont-Ferrand is involved in this perspective. Research programmes aim to improve understanding of the control of wheat grain composition during development, to permit improved and stabilized quality. Agro-physiological, transcriptional and proteomic aspects are studied in order to identify the genes involved.

The team of ABC (Adaptation du Blé aux Contraintes abiotiques) focuses on environmental factors and the distribution of nitrogen in the plant and particularly in the grain. Transcriptomic research on metaQTLs for yield, protein content and some quality traits, to localise and identify candidate genes using micro-array techniques is carried out by the PPAV (Paleogénomique des Plantes pour l'Amélioration Variétale) and DGS (Diversité Génétique et Sélection) teams. The Integrative Grain Biology (BIG) team focuses particularly on gene networks associated with storage protein synthesis. Changes in protein expression during grain formation are followed, using 2-Dimensional Electrophoresis (2-DE), and studies of their quantitative/qualitative variation and characterization. Collaboration between the teams will allow the development of a common and unique database for the soft wheat variety "Récital".

Recital is of excellent technological quality and is still much appreciated for bread-making. It is a French agronomic wheat variety registered by the Claude Benoit Seed Company in the Official French Catalogue of Species and Varieties in 1986. Récital is a winter wheat indifferent to the photoperiod, resistant to cold lodging and was revealed susceptible to several diseases like *Septoria* and *Fusarium*. It was used as parent of two mapped progenies Renan x Récital; Arche x Récital, studied in the GENOPLANTE program. The EST analysis of Récital grain was realized on eight stages of development. These ESTs are available in the NCBI DBEST public database.

A proteomic approach of the grain in development was started in 2005, under the responsibility of Dr Gérard BRANLARD on the earlier stages of grain formation and made it possible to precise the timing of protein expression for different metabolic pathways (Nadaud et al. 2010). Present research focuses on analysis of different tissues to characterize specific proteins and to improve understanding of the mechanisms involved in grain development.

The first objective of thesis was proteomic analysis of peripheral layers (PL) and of starchy endosperm (ESM) during wheat grain development; secondly we aimed at data integration of proteome and transcriptome during grain development. Peripheral layers are the inner layers of the pericarp which is made up of testa, nucellus tegument or hyaline, aleurone layer and inner pericarp. Starchy endosperm is the main reserve of starch and proteins of grain. Grains were dissected to separate peripheral layers and endosperm from developing grains, then proteomic analysis was made using 2-DE. Quantitative and qualitative (presence/absence) variations of proteins expressed during development were studied every 50°Cd in wheat grain development from one day before fertilization (0°Cd) to the end of physiological maturity (700°Cd) for peripheral layers and till complete maturity for endosperm (1006°Cd). A preliminary approach for integrating proteome and transcriptome data was proposed for carbohydrate metabolism proteins.

The present manuscript is composed of five main parts/chapters. The first part describes the bibliography related to wheat, wheat grain and its development. This part also details of recent techniques used in proteomics and transcriptomics. Proteomic analysis of wheat grain peripheral layers is described in chapter 2, and that of starchy endosperm in Chapter 3. A preliminary attempt to integrate proteome and transcriptome data is presented in part 4. Other results related to proteome comparison of peripheral layers and of starchy endosperm during development are described in part 5 together with discussions. The final part concludes the findings of the thesis and suggests future perspectives.

Chapter 1: INTRODUCTION

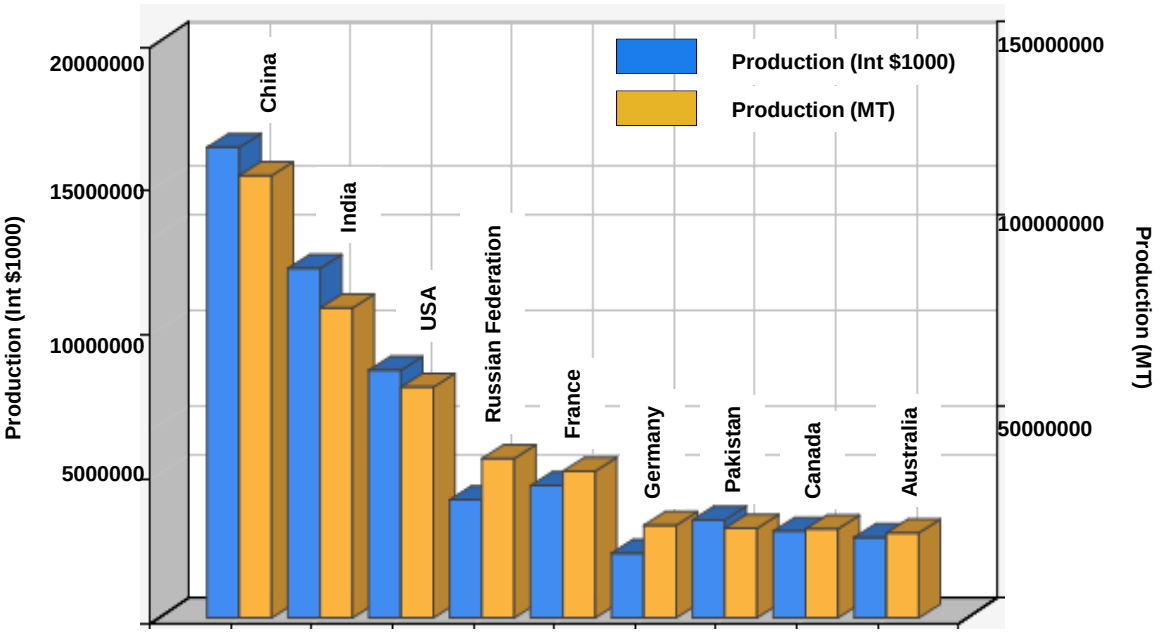


Figure 1: Top wheat producers according to FAO data 2010. Countries are presented on x-axis, while production is on y-axis.

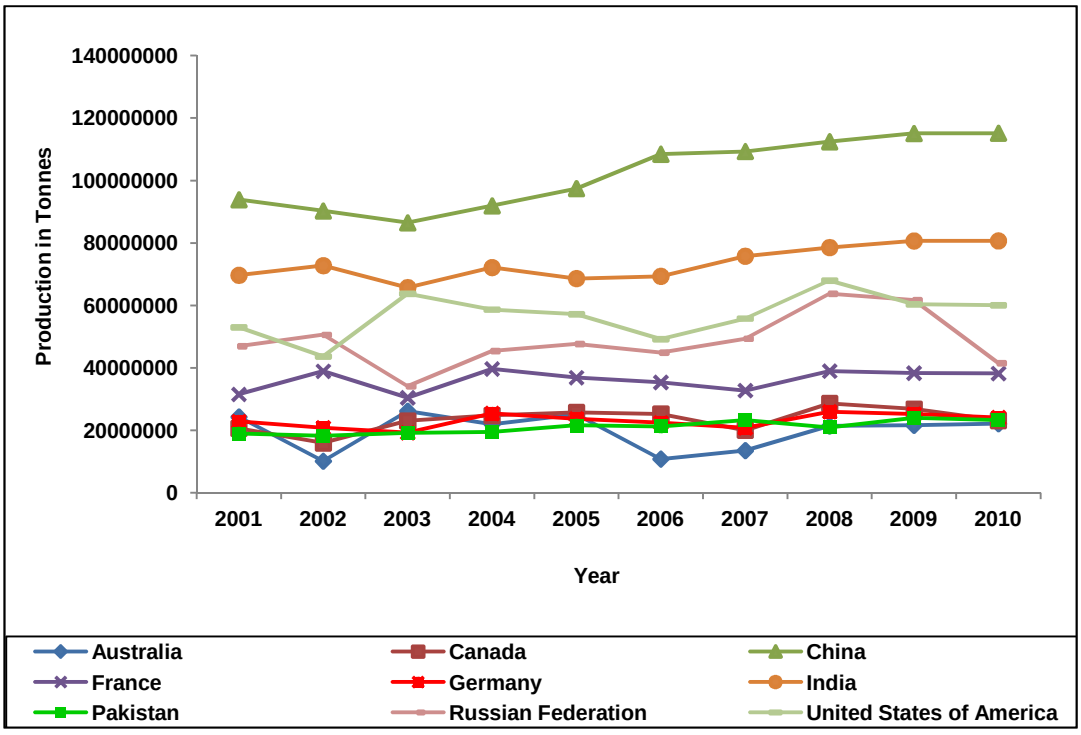


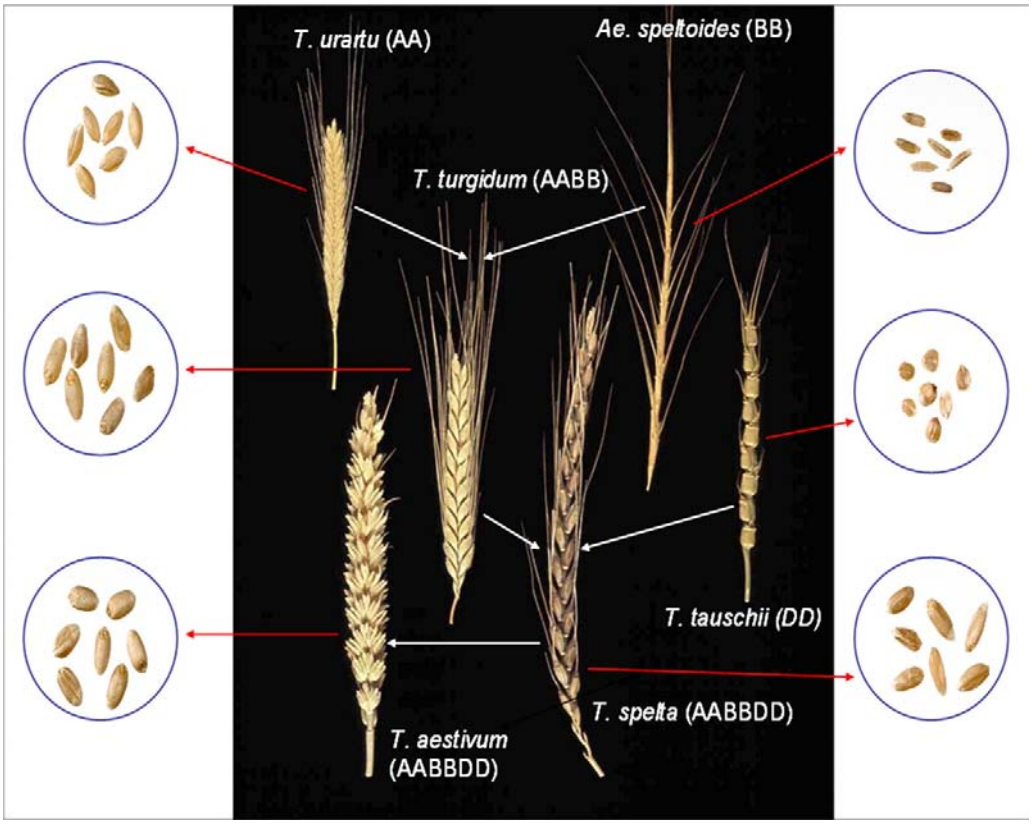
Figure 2: Ten years production data (2001-2010) of world wheat leading producers (http://faostat.fao.org/). Years of production are on horizontal axis, while production in tonnes is presented on vertical axis.

1 Wheat

Wheat has become the major source of energy and protein for human beings and occupies nearly 17% of the whole world crop area. About two-thirds of the wheat produced in the world is used for human food and about one-sixth for livestock feed. Industrial uses, seed requirements, and post-harvest losses account for the remaining withdrawals from the world wheat granaries. Wheat products are the preferred base for most cereal-based processed products, and supply ~20% of the food calories for the world population. Human wheat consumption is mainly in the form of bread, pasta, noodles, biscuits including many other confectionary products (Kumar et al. 2011).
According to FAO, its production exceeds 600 million tonnes annually with predicted production of 690 million tonnes in 2012. According to 2010, data China leads world wheat production (approximately 115 million metric tons (MMT)) followed by India and USA. France was fifth world wheat producer with an area harvested of 5426000 Ha and a production of 38.2 million tons in 2010 (http://faostat.fao.org/) (Figure1 and 2). The leading exporter was USA (21.9 MMT) while Italy was the major importer (6479090 tones) of wheat in 2009.

1.1 Origin

The “Neolithic revolution” is considered as the era when wheat originated, about 10 000 years ago. According to Vavilov 1940, wheat was first cultivated in the Mediterranean region and Southwest Asia. Later it was found that its origin was in South-east Turkey (Heun et al. 1997, Dubcovsky and Dvorak 2007). Hexaploid bread wheat was first seen around 9000 years ago with spread of cultivation to the Near East (Feldman 2001). Wheat was entered Europe via Anatolia to Greece (8000 BP before present) and then through the Balkans to the Danube (7000 BP) northwards and to Italy, France and Spain (7000 BP). It reached to the UK and Scandinavia by about 5000 BP (Feldman 2001).



Shewry, P. R. et al. 2009

Figure 3: The evolutionary and genome relationship between cultivated bread (AABBDD) and durum (AABB) wheat and related wild diploid (AA, BB, DD) are shown.

1.2 Evolutionary History

Diploid, tetraploid and hexaploid chromosome numbers of *Triticum* genus were known by 1921 (Percival 1921). The basic Triticae genome is organized into seven chromosomes which lead to the occurrence of 2n=2x=14=diploid, 2n=4x=28=tetraploid and 2n=6x=42=hexaploid wheat species. The existence of homologous and homoeologous chromosomes (genetically related chromosomes in different genomes) is useful in studies of natural phenomena of genome evolution and speciation, which were the basis of species diversity. In earlier studies it was found that *T. monococcum* and *T. turgidum* share one genome in common, while two common genomes were found between *T. turgidum* and *T. aestivum* (Sax 1922). Cytological data failed to discriminate between genomes of *T. monococcum* and *T. urartu* (Johnson and Dhaliwal 1976), as genome A donor. While molecular evidence distinguished *T. urartu* as donor of genome A for both polyploid wheats (Dvorak et al. 1993, Salse et al. 2008). *A. tauschii* was found to be the donor of D genome of bread wheat (McFadden and Sears 1946, Kihara 1944). Hence, bread wheat originated from two successive events of interspecific polyploidization (Figure 3). In a first step, *T. diccocoides* (wild ancestor of *T. turgidum*) was formed by cross between *T. monococcum* ssp urartu, (A genome donor 2n=14, AA) and an unknown specie of section Sitopsis (close to *Ae. speltoides*, B genome donor 2n=14, BB), around 0.5 to 3 million years ago. The second step occurred 7000 to 9500 years ago, through a cross between domesticated tetraploid *T. turgidum* ssp. dicoccum and the diploid *A. Tauschii* (D genome donor, 2n=14, DD) (Feuillet et al. 2008, Feldman 2001).

1.3 Wheat Genome

Wheat (*Triticum aestivum* L.) has a genome about five times larger than that of the human genome i:e ~17000 Mb and is also huge in comparison to other cereals (brachypodium 355 Mb, rice 375 Mb; barley 5500 Mb; rye 8000 Mb; maize 2365 Mb; sorghum 689 Mb) (Salse et al. 2009). It is known to be a very complex structure comprised of three independent genomes. The presence of high proportion (~80%) of repetitive sequences as compared to rice (~22%) and maize (~50%), complexify wheat genomic research as compare to other cereals (Whitelaw et al. 2003, Linkiewicz et al. 2004). However, enrichment of wheat evolutionary history data

Sub-area			Dependent area
Mediterranean (147)	Southwest Asia		Ethiopia (250)
	Anterior Asia (412)	Middle Asia (260)	
T. boeoticum (16)	*T. boeoticum* (57)		
	T. urartu (6)		
	T. araraticum (13)		
	T. dicoccoides (25)		
T. monococcum (13)	*T. monococcum* (14)		
	T. sinskajae (1)		
T. dicoccum (7)	*T. dicoccum* (15)		*T. dicoccum* (8)
	T. ispahanicum (2)		
	T. karamyschevii (3)		
	T. timopheevii (4)		
	T. militinae (2)		
	T. zhukovskyi (1)		
	T. macha (14)		
	T. vavilovii (7)		
T. spelta (14)	*T. spelta* (14)	*T. spelta* (19)	
T. durum (80)	*T.* durum (75)	*T. durum* (8)	
T. turanicum (4)	*T. turanicum* (34)	*T. turanicum* (7)	
T. turgidum (34)	*T. turgidum* (54)	*T. turgidum* (3)	
	T. carthlicum (18)	*T. jakubzineri* (1)	
T. polonicum (11)	*T. polonicum* (14)	*T. polonicum* (3)	*T. polonicum* (6)
	T. sphaerococcum (17)		
T. compactum (13)	*T. compactum* (40)	*T. compactum* (64)	
T. aestivum (25)	*T. aestivum* (59)	*T. aestivum* (142)	*T. aestivum* (33)
		T. petropavlovskyi (4)	*T. aethiopicum* (203)

Dorofeev et al. 1979

Table 1: Distribution of the species of wheat in the Ancient Mediterranean area, the origin of cultivated plants. In brackets, are the number of botanical varieties per species.

associated with cytogenetic stocks and genomic and genetic resources has resulted in significant progress in wheat genomic research (Paux et al. 2011, Philippe et al. 2012).

1.4 Wheat Classification

Seven classes of wheat were established by the Official U.S Standards for Grain (USDA-FGIS, 1987). These classes are:

Durum wheat (DU): belongs to *T. durum*

White: soft and hard white winter belongs to *T. aestivum*, club wheat *T. compactum*

Hard red spring (HRS)
Hard red winter (HRW) } belongs to *T. aestivum*
Soft red winter (SRW)

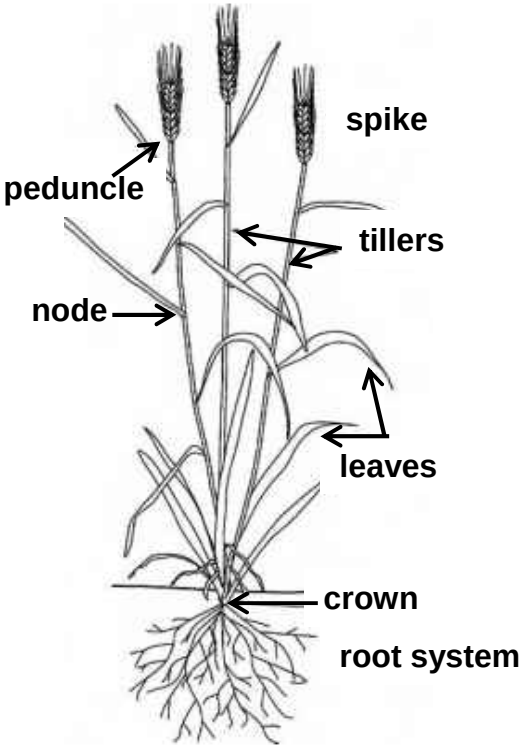
Un classed: any wheat other than white or red color

Mixed: any mixture of wheat that consists of less than 90% of one class and more than 10% of one other class or a combination of classes that meet the definition of wheat (any of various annual cereal grasses of the genus *Triticum* of the Mediterranean region and southwest Asia). Most of the European wheats (*T. aestivum*) are intermediate between red and white, and may be ranked in HRS, HRW or SRW. Soft red spring is also cultivated but on very small area.

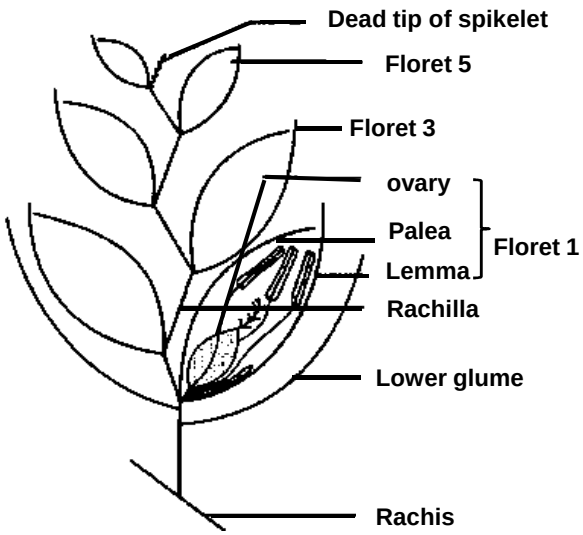
Classification into spring and winter wheat traditionally refers to crop growing seasons. Winter wheat is sown in late summer or fall. Young plants remain in vegetative phase during winter temperatures (0-5°C), and resume growth in early spring. This type of wheat is grown throughout the world where the soil does not freeze sufficiently to winter-kill the young plants. Spring wheat is generally sown in spring and is harvested in late summer than winter wheat but it can be sown in autumn in countries that experience mild winters. Its yield is generally lower than for winter wheat.

Wheat shows wide genetic variation (Table 1), many species are still in existence and currently a total of around 4000 bread wheat varieties (spring or winter) are grown around the world (Posner, 2000). Generally, four species of the genus Triticum are cultivated:

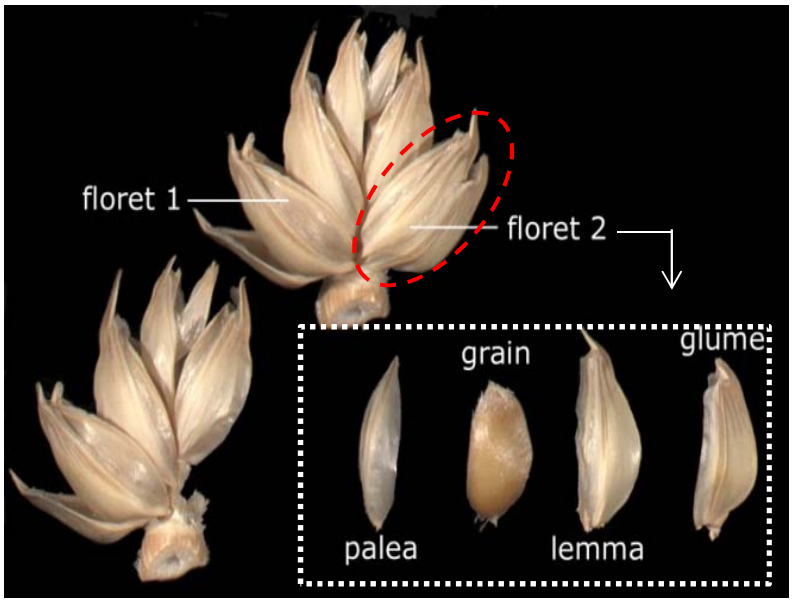
T. monococcum (diploid); *T. timopheevi* (tetraploid); *T. turgidum* (tetraploid) durum or macaroni wheats, most widely grown together with; *T. aestivum* (hexaploid) common or bread wheat. Other wheat species (*T. dicoccoides* and *T. spelta*) are still grown in some regions but in small



a: wheat plant with its different parts



b: Internal structure of wheat inflorescence. Spikelet, its florets and individual parts of florets are shown.



c: Wheat spikelet at grain maturity stage. Component parts of floret 2 have been shown. Figure modified from http://www.wheatbp.net.

Figure 4: Morphology of bread wheat plant parts (*Triticum aestivum*).

amounts. Hulled (emmer and einkorn) wheats are together called faro in Italy (Szabó and Hammer 1996, Özkan et al. 2011), while spelt continues to be grown in Alpine areas of Europe (Fossati and Ingold 2001).

Wheat crops are most successful between latitudes 30°-60°N and 27°-40°S but this cereal is unrivalled not only in its range of growing conditions (Feldman 1995) but also in its diversity.

1.5 Wheat Plant Anatomy

Botanically, the wheat grain is a single seeded fruit called a caryopsis, in which the ripened ovary wall is fused to the seed. Plant characteristics are the following (Esau 1953, Gardner et al. 1985, Soltner 1988):

Roots: Two types of roots are found, the seminal roots and the nodal roots (adventitious or crown roots), which arise from the lower nodes of the shoot.

Shoot: This is made up of a series of repeating units or phytomers, each potentially having a node, an elongated internode, a leaf, and a bud in the axil of the leaf (Figure 4a)

Tillers: They have the same basic structure as the main shoot; arise from the axils of the basal leaves.

Leave: Each leave comprises the sheath, wrapping around the subtending leaf, and a lamina (blade). At the junction of the sheath and lamina, there is a membranous structure, the ligule, and a pair of small, hairy projections, the auricles.

Inflorescence: The wheat inflorescence is called a spike or ear, where the grains are born. This is a simply branched structure in which the major axis (rachis) bears two rows of alternating secondary axes (rachilla, the axis of spikelet) (Figure 4b). The rachilla is also branched alternately, which bears a pair of empty glumes at the base and a series of florets (three to six potentially fertile) in the spikelet (Kirby and Appleyard, 1984, Acevedo et al. 2006), which are 96% self pollinated (Martin et al., 1976).The self pollination rate may vary according to environmental conditions and particularly may decrease with high temperature. Each floret consists of a pair of pales (lemma and palea) which enclose the ovary and later the caryopsis (Figure 4b and c). At the base of the floret, there is a pair of lodicules that swell at the time of fertilization, pushing the pales open and allowing the anther on elongated filaments to emerge.

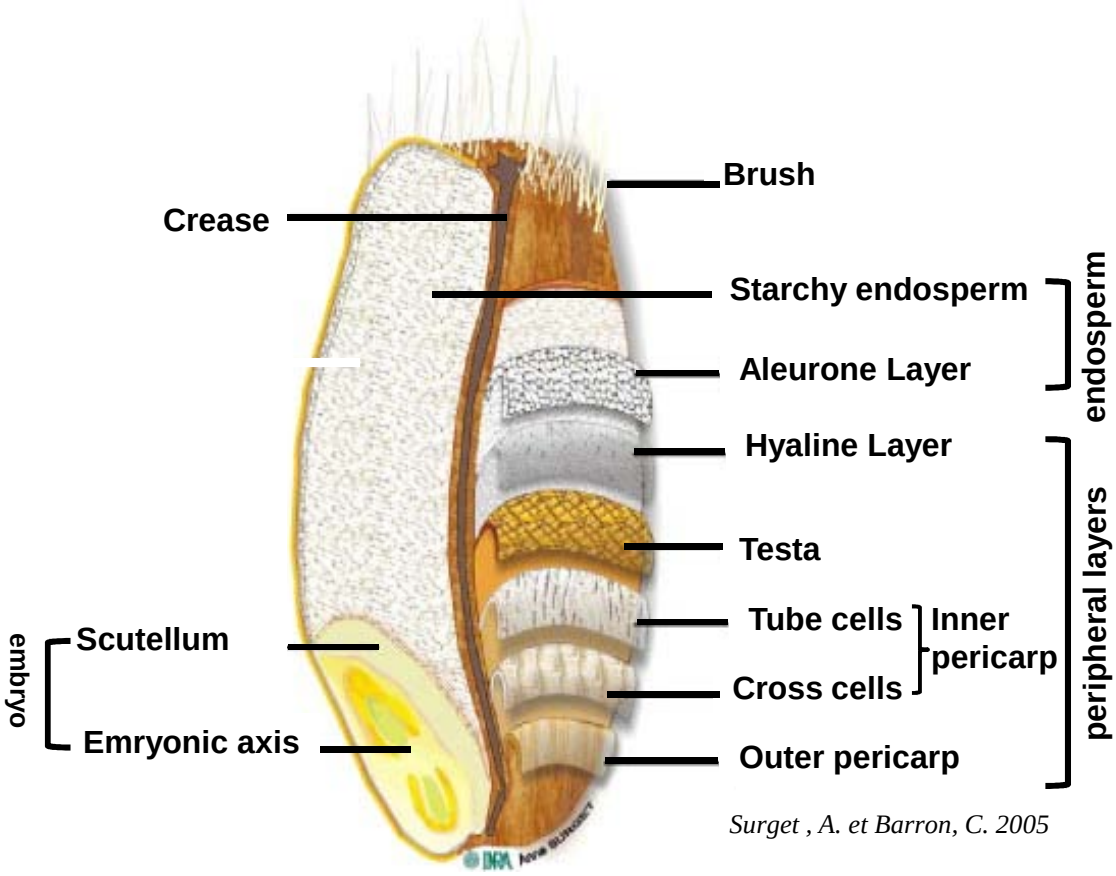


Figure 5: Longitudinal cut of wheat kernel. Three major parts of grain are shown: peripheral layers, endosperm and embryo.

	AX g/100g **b**	Glc **a**	A/X	Bound FA g/100g **b**	DHD	Bound FA + DHD g/100g AX **c**
Outer Pericarp	47.0	25.9	1.16	0.33	0.23	1.2
Inner Pericarp + Testa + NE	38.9	15.7	0.42	0.58	0.06	1.4
Nucellar epidermis (NE)	60.0	4.4	0.12	0.48	0.06	1.8
Aleurone	24.4	11.8	0.44	0.81	0.03	3.5
Starchy endosperm	1.6	76.2	0.84	0.005	traces	0.3
Embryonic axis	14.8	3.0	1.35	0.04	0.01	0.4

Saulnier, L., 2012

Table 2: Polysaccharide composition of tissues in mature wheat grain.
a: mainly present as mixed linked beta-glucan in aleurone, starch in starchy endosperm and cellulose in other tissues; **b:** in g/100g of tissue; **c:** Ferulic acid+ DHD in g/100g of AX.
AX : sum of arabinose and xylose ; Glc : glucose.A/X : arabinose/xylose ratio ; FA : Ferulic acid ; DHD sum of ferulic acid dehydrodimers.

1.6 Wheat Grain Structure

Wheat grains differ in their shape from almost spherical to long, narrow or flattened, but in general they are oval shape, 5-9 mm length and of weight 35-50 mg. They have a crease down one side which indicates the point of connection to the flower (Belderok et al., 2000). The wheat grain is composed of mainly three parts (Figure 5).

Peripheral Layers (PL): hyaline layer, testa, inner and outer pericarp
Endosperm (ESM): starchy endosperm and aleurone layer (AL)
Embryo: scutellum and embryonic axis

1.6.1 Peripheral Layers (PL)

The outer layers of wheat grain protect the embryo and endosperm part of the grain. To protect these parts, more than half the bran (PL + aleurone layer) consists of water-insoluble fiber components (53%) (Table 2).
Wheat bran fiber contains, essentially, cellulose and pentosans, polymers based on xylose and arabinose, which are tightly bound to proteins and their ratios, differ between these layers (Antoine et al. 2002, Antoine et al. 2003, Parker et al. 2005, Šramková Z 2009). Particle size distribution, microstructure, starch content and chemical composition also differ between bran samples, studied on both wheat and rye bran (Kamal-Eldin A 2009). Barron et al. 2007 made an attempt to identify the possible tissue specific biochemical markers of wheat. The pericarp (ovary/fruit wall), is the ripened ovary wall which is dead at grain maturity. Most of the tissues are devoid of cytoplasm and have lignified walls. Its composition include: 30-40% heteroxylans, 25-30% cellulose, ~10 lignins and 5-7% proteins (Pomeranz 1988, Fincher and Stone 1986).

Outer pericarp is composed of outer epidermis, hypodermis and thin walled starch-filled parenchyma. The epidermis is 15-20 µm thick and is a complete layer without intracellular space which covers the entire grain surface except for the rachilla. Cell dimension and shape vary according to cultivar. Study of Bradbury et al 1965a showed that it is formed of long narrow cells (80-300 µm length, 25-48 µm width, 3-9.5 µm thickness) that are alternately arranged. The

mature hypodermis lies just below the epidermis, to which it resembles in structure and appearance at maturity.

Inner pericarp is composed of cross and tube cells, found only in grasses (Rost and Lestern 1973). These elongated cells (mainly cross cells) are arranged perpendicular to epiderm cells, and are aligned in palisade. They are photosynthetic (chlorenchyma) in nature, and give the green colour to developing caryopsis in early stages (Morrison 1976, Evers and Millar 2002). Their size is between 100-150 µm X 10-15 µm. The tubes cells are elongated cells disposed in quincunx and above all present on dorsal part of grain. Their size is 120-250 µm X 12-15 µm, 5-10 µm wide.

1.6.1.1 *Testa (seed cover)*

In the wheat grain it is fused with the fruit coat (pericarp), although they are united physically but botanically can be distinguished. This cell layer develops from the integuments that surround the nucellus at the time of fertilization. It has cells that are 100-191 µm long, 9-20 µm wide and 5-8 µm thick (Bradbury et al 1965a).

1.6.1.2 *Hyaline layer (or nucellar layer)*

This cell layer is located between the endosperm and the seed coat in the mature caryopsis, an amorphous layer is present to separate this layer from aleurone layer of endosperm. The elongated cells of this layer vary widely in size (30-200 µm long, 15-40 µm wide, 12-20 µm thick). At fertilization, this layer is present as a thick tissue and it degenerates in parallel to grain maturity (Baldo et al. 1983). Degeneration of this layer provides nutrients and space to the developing endosperm and embryo and may also influence further changes in other grain tissues.

1.6.2 Endosperm (ESM)

The endosperm is a triploid tissue, a consequence of double fertilization (biological process in which one sperm nucleus fuses with the egg to produce the embryo, while a second sperm nucleus fuses with two polar nuclei to form the triploid endosperm). The origin of this tissue is also intrinsically linked to this unique biological process (Lopes and Larkins 1993).

Constituents (µg/g)	Pericarp	Aleurone layer	Albumen
Iron	97	338	10
Manganese	56	78	5.5
Total Phosphorus (%)	0.056	3.17	0.08
Thiamine	0.6	16.5	0.42
Niacine	20-25.7	613-741	8.79
Pyridoxine	6	36	-
Riboflavine	Traces	10	0.28

Pomeranz, 1988

Table 3: Principal minerals and amino acids content of wheat kernel tissues.

Grain Contents	Whole grain	Mealy endosperm	Bran	Embryo
Proteins	16	13	16	22
Fats	2	1.5	5	7
Carbohydrates	68	82	16	40
Dietary fibers	11	1.5	53	25
Minerals (ash)	1.8	0.5	7.2	4.5
Other component	1.2	1.5	2.8	1.5
Total	100	100	100	100

Table 4: Caption of The wheat grain: Chemical composition of the whole wheat grain and its various parts (converted to percentages on a dry matter basis)

1.6.2.1 ***Aleurone Layer***

The outer endosperm, the aleurone layer (AL), has a special structure: it consists of single layer of cubic shaped cells whose cell wall layers are rich in AX and inside the cavity are proteins and other enzymes. It is a tissue rich in proteins, lipids, phytates, vitamins (B1, B2, B3, B6, B9 and E) and minerals (P, K, Mg and Fe) (Table 3) (Pomeranz 1988, Brouns et al. 2012). These proteins and enzymes play a vital role in germination processes, being the only living tissue of mature grain along with embryo. AL is modified where it borders the endosperm cavity in the region of the crease (Winton et al. 1932). Modified AL is found to be involved in formation of the crease (Evers 1970).

1.6.2.2 ***Starchy endosperm***

The inner endosperm is referred to as mealy or starchy endosperm and as the name indicates, contains the totality of energy-yielding starch of the grain, which is used by embryo as the major source of nutrition for development (Belderok 2000). In contrast to its high concentration of carbohydrates, this tissue is poor in proteins, fats, dietary fibers and minerals as compared to embryo and AL (Table 4). According to Popineau and Pineau 1988, the distribution of proteins and weight is not equal among different grain tissues, as can be seen in the Table 8 on page 25. In starchy endosperm three types of cell are reported, differing according to their size, shape and sites of occurrence (Greer et al 1951).

i) Peripheral or sub-aleurone cells are adjacent to the aleurone cells. They are of similar size to these cells with about 60 µm of diameter. They are rich in AX and lenticular starch granules of A type (diameter >10 µm). **ii) Prismatic cells** radiate in columns from their junction with sub-aleurone cells. These cells are 128-200 µm long with 40-60 µm wide. **iii) Central cells**, are generally round or polygonal and have a length of 72-144 µm and a width of 69-120 µm (Bradbury et al. 1956b, Evers and Millar 2002). All three types of starch granules (A, B, C) are present in these two types of cells (prismatic and central), which differ in size and shape. The large one is lens shaped (A type) and the small one is near-spherical (B and C type).

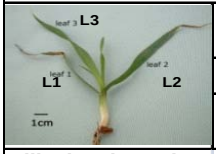
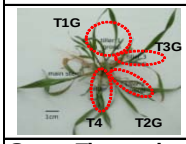
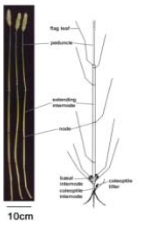
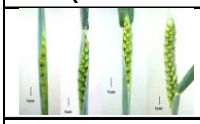
Stage (DAS)	General Description	Feekes Scale	Zadoks Scale	Additional Comments
Germination (0-42)				
	Dry seed		0	
	Imbibition complete		3	Seed typically at 35 to 40% moisture.
	Leaf just at coleoptile tip		9	
Seedling Growth (until ~77)				
	First leaf through coleoptile	1	10	
	First leaf unfolded		11	
	9 or more leaves unfolded		19	
Tillering (78-89)				
	Main shoot and 1 tiller	2	21	
	Main shoot and 3 tillers		23	Many plants will only have 2 or 3 tillers per plant at recommended populations.
	Main shoot and 6 tillers	3	26	Leaves often twisting spirally.
Stem Elongation(90-101)				
	Pseudostem erection	4 and 5	30	
	1st detectable node	6	31	Jointing stage
	2nd detectable node	7	32	
	4th detectable node		34	Only 4 nodes may develop in modern varieties.
	Flag leaf visible	8	37	
	Flag leaf ligule and collar visible	9	39	
Booting (102-113)				
	Flag leaf sheath extending		41	Early boot stage.
	Boot swollen	10	45	
	Flag leaf sheath opening		47	
	First visible awns		49	In awned varieties only.
Head (Inflorescence) Emergence				
	First spikelet of head visible	10.1	50	
	Head completely emerged	10.5	58	
Pollination (Anthesis) (up to ~128)				
		10.51	60	Flowering usually begins in middle of head.
		10.52		Flowering completed at top of head.
	Beginning of flowering	10.53		Flowering completed at bottom of head.
	Flowering completed		68	
Grain development (129-190)				
	Kernel (caryopsis) watery ripe	10.54	71	
	Early milk		73	
	Medium Milk	11.1	75	Milky ripe.
Milk Development (129-143)	Late Milk		77	Noticeable increase in solids of liquid endosperm when crushing the kernel between fingers
	Early dough		83	
Dough Development (144-158)	Soft dough	11.2	85	Mealy ripe: kernels soft but dry.
	Hard dough		87	
Ripening (159-190 until harvest stages)	Kernel hard (hard to split by thumbnail)	11.3	91	Physiological maturity. No more dry matter accumulation.
	Kernel hard (cannot split by thumbnail)	11.4	92	Ripe for harvest. Straw dead.

Table 5: Wheat growth stages modified from (http://www.uky.edu). Pictures and data about DAS is taken from http://www.wheatbp.net. L=leave, T=Tiller, G=group, DAS=Days after sowing.

1.6.3 Embryo

The embryo is situated on the front dorsal side of grain and consists of scutellum (massive storage organ, considered as a cotyledon) and germ cells that gives the seedling after germination. It is rich in proteins, lipids and minerals (Table 4). Wheat germ is available as a separate entity because it is an important source of vitamin E. It has only half the glutamine and proline of flour, but levels of alanine, arginine, asparagine, glycine, lysine and threonine are double (Evers and Millar 2002, Cornell 2003).
At the time of grain germination, the embryo was found to be involved for absorption of glucose from the endosperm (Aoki et al. 2006).

2 Development of Wheat Plant and Grain

Winter wheats, like many other cereal grains, require vernalization (Dennis and Peacock 2009). They tolerate low temperatures for a specific required length of time to become competent to flower (plants convert from vegetative to reproductive growth). This process prevents the damage of the cold-sensitive flowering meristem during the winter. There is no requirement of vernalization in spring wheat varieties (Acevedo et al 2006). VNR1 gene is critical in division of polyploidy wheat varieties into the winter and spring categories (Tranquilli and Dubcovsky 2000, Law and Giorgi 1975). Vernalization may occur at three stages of the growing cycle of the wheat plant: during germination, during vegetative plant growth and during seed formation in the mother plant (Flood and Halloran 1986).

Five different scales are used worldwide for wheat growth stages. This classification varies from simple division in three phases of development to construction of detailed scales which covers all visual steps of development. Commonly used scales are Zadoks and Feekes-Large scales. The comparison between the two scales is not too easy but some key stages of development allow us to align these two scales (Table 5). In Feekes scale developmental stages are designated on a scale of 1 (seedling growth) to 11 (ripening) (Large 1954). The Zadoks scale is much more descriptive. Globally it goes from primary stage 00 (dry seed) to 90 (ripening) (i:e 10 primary stages, each of which is divided into 10 secondary stages = total of 100 stages) (Zadoks et al.

1974). The time span of each development phase depends essentially on genotype, temperature, day length and sowing date (Acevedo et al. 2006).

2.1 Germination to the End of vegetative growth

Germination is a complex process during which the imbibed mature seed must quickly shift from a maturation to a germination driven program of development and prepare for seedling growth (Nonogaki et al. 2010). Phytohormones such as abscisic acid (ABA), gibberellins (GAs), ethylene, brassinosteroids, salicylic acid, cytokinin, auxin, jasmonic acid and oxylipins are one the important factors influencing this phenomenon by forming an interlocked signaling network (Rajjou et al. 2012). ABA and gibberellins have antagonistic effects, where ABA is present at later stage of grain development to inhibit germination and thereafter grain imbibition its level is reduced, while GAs are essential germination activators. ABA/GA ratio regulates the metabolic transition required for germination process (Liu et al. 2010).

The mature grain contains enough stored reserves in the endosperm that are sufficient for the growth of embryo wheat plant. Minimum water content required in the grain for wheat germination is 35 to 45% by weight (Evans et al. 1975) with temperature range of 4-37°C being optimal from 12-25°C. Wheat crop emergence occurs at 3-4 leaf primordial stage of seed embryo and almost half of the leaf primordia already initiated (Baker and Gallagher 1983a, 1983b, Hay, and Kirby 1991). The first roots that appear are the seminal roots followed by coleoptiles which protect the emergence of the first leaf. Germination process lasts around 0-42 DAS (days after sowing of seeds, Table 5) and vegetative stage may vary from 60 to 150 days.

With the emergence of the first leaf seedling, growth begins and continues until the next stage (tillering). Not all tillers produce spikes in wheat and many of them abort before anthesis (Gallagher and Biscoe 1978). Next is the stem elongation (jointing) phase which occurs as a result of internodes elongation. This phase completes with the emergence of last leaf from the whorl (flag leaf) (Krumm et al. 1990). Shortly after flag leaf emergence, the flag leaf sheath and the peduncle pushed up the developing head through the flag leaf sheath which obtains a swollen appearance to form a boot. Boot stage is rather short and coincides with meiosis which starts in the middle of the spike continuing later, above and below this zone, in wheat and barley plant growth (Zadoks et al. 1974).

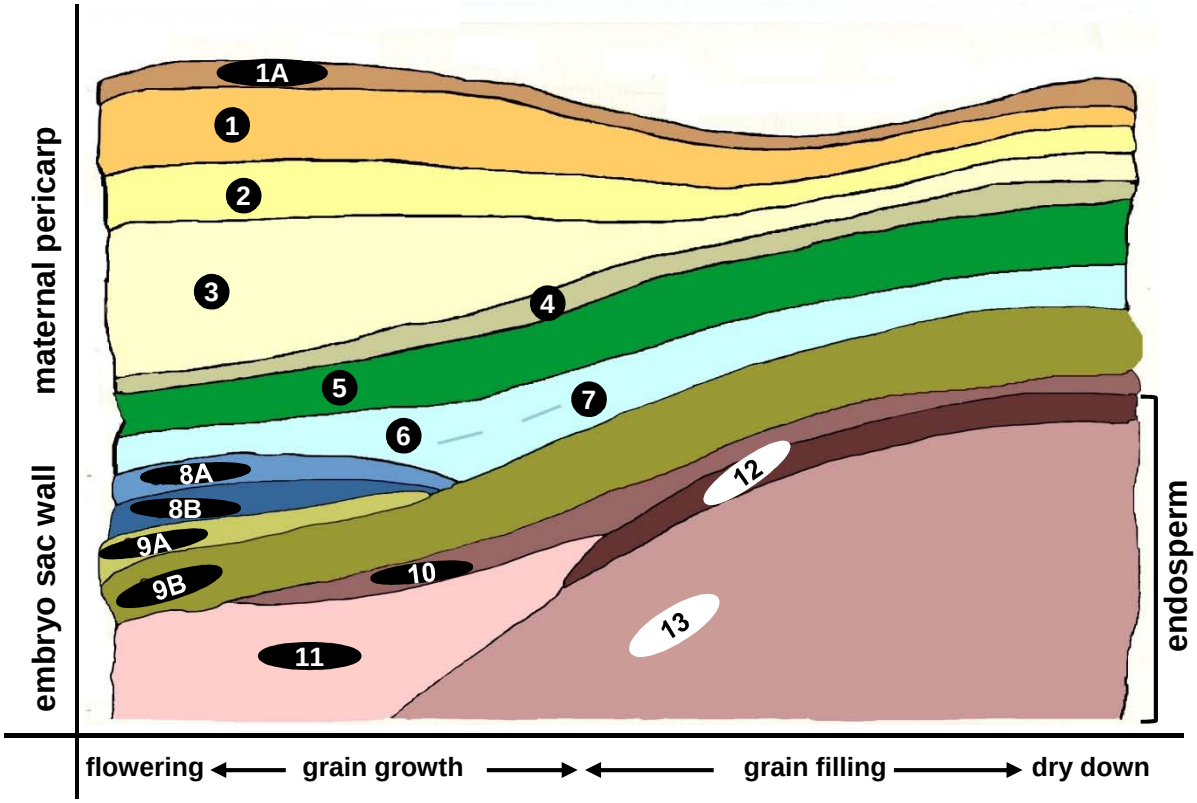


Figure 6: Schematic representation of theoretical changes from flowering to grain maturity. Each number corresponds to cell layer which are named in Table 6. (modified from http://www.wheatbp.net)

	Pericarp	**Number in diagram**	**Name of tissue**	**Becomes later**	**Collectively called / Mill Fractions**	
PL		1A	Cuticle of Outer epidermis		Beeswing	Bran
		1	Outer epidermis	Outer pericarp		
		2	Hypodermis			
		3	Parenchyma (thin-walled)			
		4	Intermediate cells	Inner pericarp		
		5	Cross cells			
		6 becomes7	Inner epidermis/ Tube cells			
	Embryo sac wall	8A and 8B	Outer integument	Pigmented seed coat (testa)		
		9A and 9B	Inner integument			
		10	Nucellar epidermis			
		11	Nucellus	degenerates		
ESM		12	Aleurone		Removed with bran	
		13	Starchy endosperm		White flour	
Germ		Not Shown	Embryo and scutellum		Germ	

Table 6: Main constituents of grain and their corresponding cell layers, along with general terms used by farmers and millers for each fraction . (modified from http://www.wheatbp.net)

2.2 Flowering and fertilization

Heading or inflorescence emergence stage begins (Feekes 10.1; Zadoks 50) when the tip of the spike (head) can be seen emerging from the flag leaf sheath, and continues until the complete emergence of head (Feekes 10.5; Zadoks 58) which than enters in flowering stage (anthesis) (Table 5). This is the transitional stage from vegetative to initiation of reproductive phase. Like meiosis, anthesis also normally begins in the center of the head and progresses to the top and bottom of the head during 3-5 day period (Peterson 1965). All heads on a plant pollinate within a few days (starts from the main stem with tillers slightly later). Pollen formation and pollination are very sensitive to environmental conditions. Double fertilization event takes place in carpel, within the embryo sac.

2.3 Grain Development

Important events during formation of wheat kernel can be divided in three main phases of development i.e Grain growth, Grain filling and Grain dehydration.

2.3.1 Grain Growth

After double fertilization, an overview of endosperm and embryonic sac wall changes is presented in Figure 6 and Table 6. During this phase grain parts undergoes following changes:

Peripheral Layers (PL)

A network of complex changes in these layers lead to degeneration of outer integument and nucellus that were present at start of grain development (Figure 7a and 7b). In embryonic sac walls only inner integument is viable throughout grain development (Figure 6).

Endosperm (ESM)

ESM development in cereals has been studied widely (Olsen et al. 1999, Olsen 2001, 2004, Sabelli and Larkins 2009, Xu et al. 2010, Opanowicz et al. 2011). ESM development can be

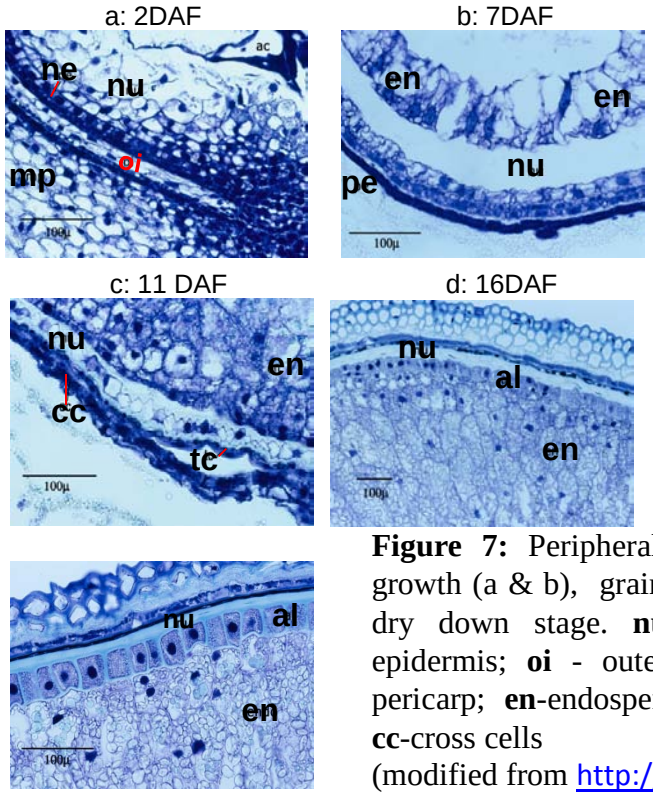
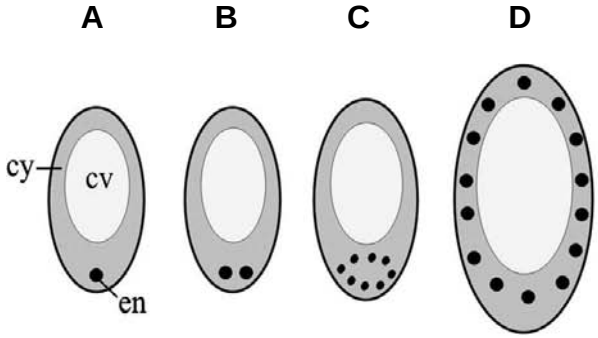
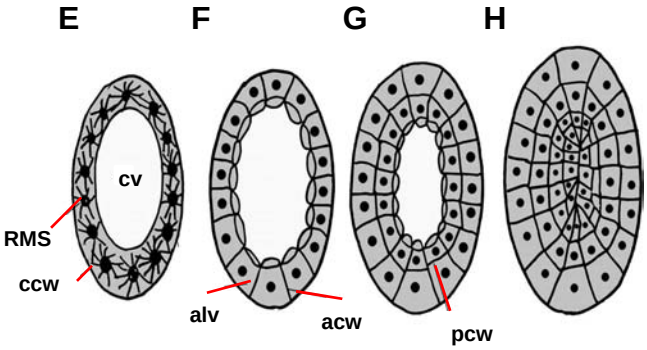


Figure 7: Peripheral layers changes during grain growth (a & b), grain filling (c & d) and just before dry down stage. **nu** – nucellus; **ne** - nucellar epidermis; **oi** - outer integument; **mp** - maternal pericarp; **en**-endosperm; **pe**-pericarp; **tc**-tube cells; **cc**-cross cells
(modified from http://www.wheatbp.net)



The Endosperm Coenocyte

(A) central vacuole (cv) fills up most of the volume. endosperm (en), cytoplasm (cy). **(B)** central cell nucleus division. **(C)** eight endosperm nuclei stage **(D)** complete endosperm coenocyte



Cellularization of the coenocytic endosperm

Olsen, O.A., 2004

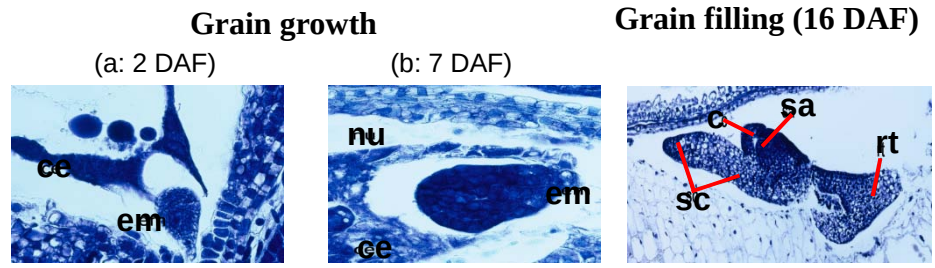
(E) Radial Microtubal System (RMS) form on nuclear membranes in the cereal endosperm coenocyte. ccw, central cell wall; cv, central vacuole. **(F)** alveoli (alv) formation. acw, anticlinical cell walls **(G)** periclinal cell wall (pcw) formation **(H)** endosperm at complete cellularization stage

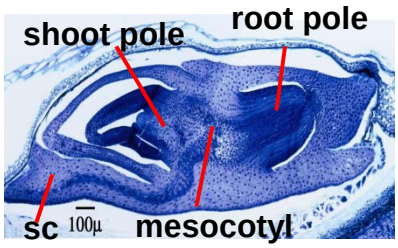
Figure 8: Endosperm development in cereals, coenytic and cellularization stages.

divided in four phases: coenocytic, cellularization, accumulation and dehydration (Olsen 2001). First two events are characteristic of grain growth period. In wheat the cell division is accompanied with water accumulation in the young grain.

Coenocytic ESM: The mitotic activity of the triploid nucleus results in a coenocytic endosperm (4-5 DAP) where the cell wall material is formed by microtubules that are condensed near the nucleus membrane (Olsen 2001). In Figure 8, triploid endosperm nucleus (en) is located in the basal cytoplasm of the central cell. A large central vacuole (cv) fills up most of the volume, surrounded by a thin line of cytoplasm (cy) (**A**). Next is the division of central cell nucleus without the formation of a functional interzonal phragmoplast, and no cell wall is formed between sister nuclei (**B**). Eight endosperm nuclei are located in a single plane in the basal endosperm coenocytes, after three rounds of nuclear division (**C**). At the completion of this phase, nuclei are evenly spaced in the entire peripheral cytoplasm (**D**) (Olsen 2004). More than 2 000 syncitial nuclei are reported in several Triticum species, as compared to 512 in maize (Bennett et al. 1975, Sabelli and Larkins 2009). The number of cells in teguments is determined early, approximately 5 DAP (Schnyder and Baum 1992).

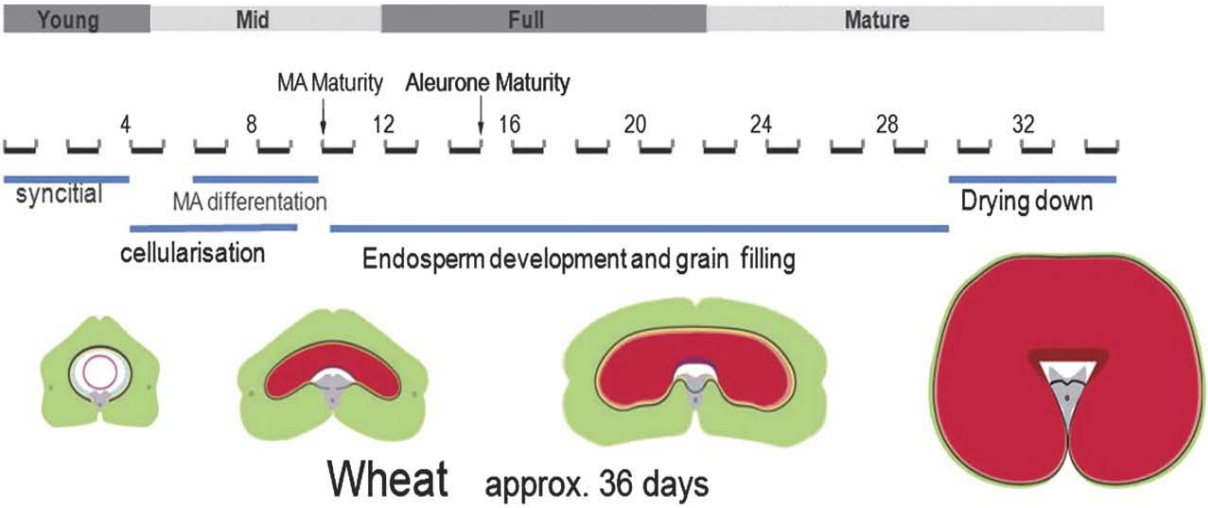
Cellularization: After the nuclei migrate to the periphery of the central vacuole, the cell walls are formed. Centripetal cell divisions occur first in anticline mode to form anticlinal cell walls (acw) which form tubes or alveoli (alv) to surround nucleus with their open ends toward the central vacuole (Figure 8 E and F). The division in pericline mode results in periclinal cell walls (pcw) that separates the outer layer of cells from a new layer of alveoli. (Figure 8 G). This mode of division continues in the innermost layer of alveoli until complete cellularization of endosperm (Figure 8 H) (Olsen 2001, 2004). AL appears just before the start of grain filling and is clearly identifiable in cereals between 6-10 DAP, as can be seen for wheat in Figure 6. In wheat single layer of AL is found like maize, while in barley it consists of up to three layers (Jestin et al. 2008, Sabelli and Larkins 2009). The number of cells at the end of the cell division was reported to be linked to final kernel mass (Brocklehurst 1977). Final kernel volume appears to be established at the end of cell division. These observations led physiologists to state that the potential size of the wheat kernel is determined early (9 to 12 DAP).





Before dry down (26DAF)

Figure 9: Development of embryo, a & b events are in first phase of grain development (grain growth phase). Embryo is fully differentiated with embryonic shoot and embryonic root. **ce**=cellular endosperm; **em**=embryo; **nu**=nucellus; **sc**- scutellum; **rt**- root tip; **c** - coleoptile; **sa** - shoot apex. (modified from http://www.wheatbp.net)



Modified from Opanowicz, M. et al. 2011

Figure 10: Wheat grain development focusing on the key stages in endosperm development. Grain tissues are indicated by colour: pericarp, green; integuments; black; nucellar projection, grey; central endosperm, red. MA is modified aleurone.

Embryo

In newly formed zygote cell divisions are slow. A basal cell produced as result of one transverse division and it does not divide again. A five celled globular shaped embryo is formed in consequence of apical cell divisions. Seven days after fertilization, cell division continues but differentiation is slow, so that there is a simple dermatogen surrounding a central core of cells. (Figure 9 a and b).

2.3.2 Grain Filling

This stage corresponds to the growth of endosperm cells when the storage components, starch and proteins, are accumulated.

Peripheral Layers (PL)

These layers were differentiated at grain growth stage. In grain filling stages changes in nucellus occur. At 16 DAF it becomes progressively compressed as the endosperm packs with starch. (Figure 7d).The pericarp tissue was the major site of starch deposition until 3 DAP, then after 6 DAP, size and the number of granules in endosperm gradually increased (Li et al. 2012).

Endosperm (ESM)

This phase is accumulation stage of ESM, with exponential synthesis and accumulation of storage proteins in protein bodies. A-type starch granules are also synthesized and starch granules of B-type are initiated and accumulate at 26 DAF (Simmonds and O'Brien 1981). The dorsal cells of the AL continue cell division, helping kernel expansion (Altenbach et al. 2003) (Figure 10). When the starch component is accumulated, there is a change in the colour of the kernel from green to yellow. Transcriptomics analysis of kernel development revealed that the genes for protein and starch decreased and genes associated with plant defense are particularly expressed during the end of grain filling (Vensel et al. 2005, McIntosh et al. 2007).

At the end of grain filling the cells of AL are embedded in a firm matrix and have a granular appearance as the vacuoles are filled with proteins (Figure 7).

Embryo

The embryo continues to divide and increase in size, as a result by 11 DAF a cleft is seen on the dorsal side. Rapid differentiation of the two components of the embryo occurs at this stage; the embryonic axis (comprising the mesocotyl and the shoot & root poles) and the scutellum. At 16 DAF, scutellum becomes a fully differentiated organ that transfers nutrients from the endosperm reserves to the embryo for its own further development.

2.3.3 Grain dehydration-maturation stage

Peripheral Layers (PL)

Peripheral layers undergo the process of programmed cell death at this stage and are yellow in color.

Endosperm (ESM)

This stage is characterized by a decrease in water content. When the water content reaches 55% dry matter (DM), the starch and protein accumulation process is stopped and the dry matter content is at the maximum. This stage is also called physiological maturity, usually at this time the flag leaf and spikes turn yellow (Hanft and Wych 1982). In durum wheat the grain filling was reported to be stopped at 45.9% of grain moisture (Ferreira et al. 2011). At maturity, the water content is between 10 to 15% of DM (Young and Gallie 2000).

Embryo

Although during the grain filling period the structure of the embryo is completed, but it continues to receive storage reserves of lipid droplets and protein bodies until later. Within the space of forty days a perfectly formed miniature plant (with fully differentiated embryonic shoot and embryonic root) is developed from a single cell of the fertilized ovum.

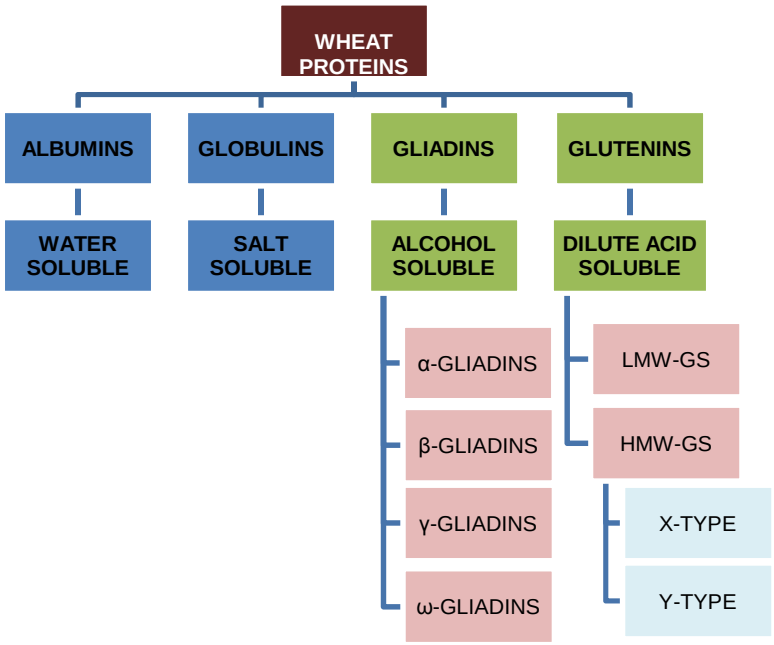


Figure 11: Wheat protein classification according to their solubility.

	Soluble proteins (Non-Prolamins)		Insoluble proteins (Prolamins)	
Protein name	albumins	globulins	gliadins	glutenins
% in wheat proteins	15-20%		35%	45%
Natural function	Proteins for grain metabolism		Reserve proteins for future plant	
Molecular weight	5 -90 kDa		25-75 kDa	100- 1.10^3 kDa
Lysine	3.2	5.9	0.5	1.5
Threonine	3.1	3.3	1.5	2.4

Osborn 1907

Table 7: Wheat proteins and their characeristics.

3 Wheat Grain Contents (Composition)

3.1 Non Protein Contents

The wheat grain (*Triticum aestivum*) is composed of 12-18% water, 63-74.5% glucids (starch and sugar), 8-12% proteins, 1.5-2% of lipids, 2.5-3% cellulose and 1.5-2% minerals. Wheat kernel components, particularly storage proteins and starch, have received a great deal of attention in the last two decades. Distribution of these contents is presented in (Table 2, 3 and 4). In this section, wheat proteins will be presented in detail.

3.2 Protein Contents

Traditionally wheat proteins have been classified into four types according to their solubility (Figure 11), soluble; metabolic proteins (albumins and globulins) and insoluble, gluten or storage proteins (gliadins and glutenins) (Osborne 1907, Table 7).

3.2.1 Metabolic Proteins (Non-prolamins or Soluble)

Albumins and globulins of flour are also known as soluble, non-prolamins or metabolic proteins. They are not specific to the endosperm; they can easily be extracted from any kernel tissue (endosperm, aleurone layer, and embryo) (Table 8).They are composed of thousands of enzymes and proteins needed for cell survival (most albumins) and cell structure (many globulins). They represent 15-20% of total wheat proteins (Shewry et al. 1986). Some high molecular weight albumins and globulins of triticins can also serve as storage proteins (Gupta et al. 1991, Singh and Skerritt 2001). Genes of globulins were localized mainly on chromosomes 4D, 5D, 6D, 7B and 7D, while characterization of products of these genes by IEF revealed that majority of them have pI that vary from 5.50 to 7.89 (Singh and Skerritt 2001). More information regarding this category of proteins is provided in Chapter 3.

3.2.2 Storage Proteins (Prolamins or non-soluble)

Prolamins are defined as proteins extractable in aqueous ethanol. Storage proteins (gliadins and glutenins) because they are very rich in proline and glutamine are also called as prolamins. As in

Grain part	Weight (% of total grain)	Protein (%Dry Mass)	Protein type
Embryo	3	35-40	Albumins and globulins
Aleurone layer	7-9	30-35	Albumins and globulins
Pericarp	4	6-7	Albumins and globulins
Starchy endosperm	80	9-14	Albumins and globulins/ Prolamins

Popineau and Pineau 1988

Table 8: Grain parts with their corresponding weight, proteins and protein types.

addition to gliadins, the low molecular weight glutenins are also soluble in aqueous ethanol after reduction. They are evolutionarily closely related to the rest of the wheat prolamins (Shewry and Tatham 1999).

The storage proteins of wheat are unique because they are also functional proteins. They do not have enzymatic activity, but they are the only cereal proteins to form strong, cohesive dough. Gluten, which is simply obtained by water-washing of flour, is widely known to be responsible for its mixing properties. This water-insoluble network is a complex physico-chemical system composed of polymeric glutenins and monomeric gliadins (Bietz and Wall 1972).

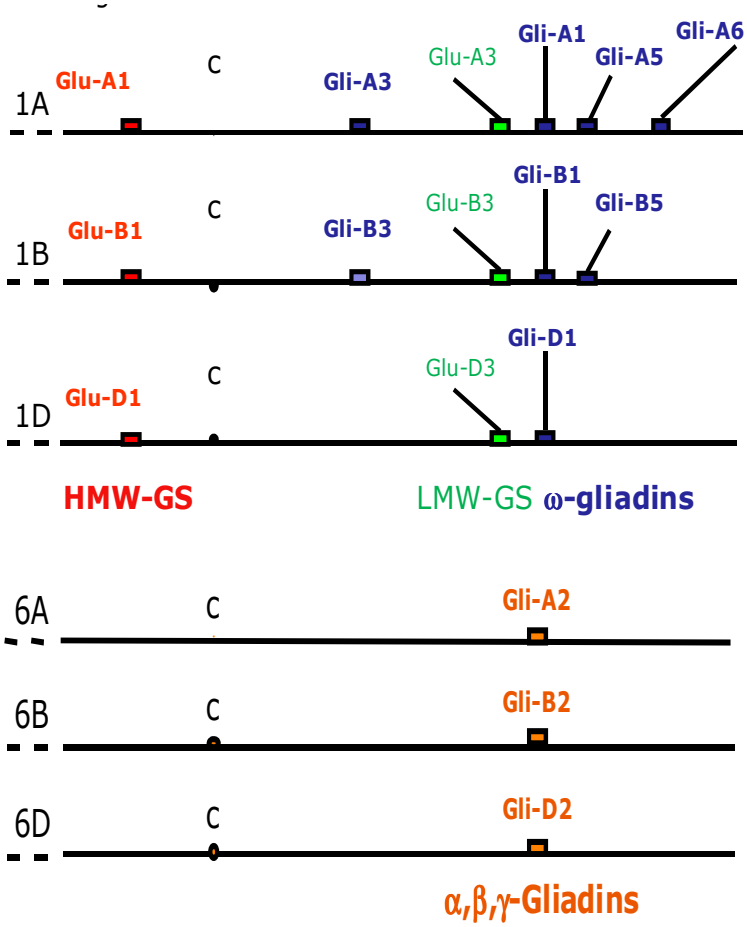
3.2.2.1 ***Glutenins***

They are known to provide elasticity whereas gliadins mainly influence the extensibility and viscosity of the dough system (Bietz and Wall 1980, Clare Mills et al. 1990). Glutenins are classified according to their electrophoretic mobility in two groups HMW-GS (High Molecular Weight Glutenin Subunits) and LMW-GS (Low Molecular Weight Glutenin Subunits). HMW-GS (72 -130 KDa) and LMW-GS (30-60 KDa) are soluble in acid and dilute base respectively.

3.2.2.2 ***Gliadins***

This category is divided into four groups: alpha, beta, gamma and omega, according to their electrophoretic mobility. They represent approximately 50% of gluten proteins and are often considered as merely acting as diluants (Macritchie 1992). Gliadins are often resolved using A-PAGE (Acidic Poly Acrylamide Gel Electrophoresis) for purposes such as allele identification, variety identification and seed purity controls. Gliadins and glutenins can also be separated using SDS-PAGE. This technique was largely employed to study the genetic determination of these proteins (gene chromosomal location, allelism) in the 1990's.

Wheat gluten proteins are secretory in nature i:e their synthesis, folding and deposition take place within the endomembrane system of the plant cell. Two different pathways were observed for their secretion and accumulation. They may pass from ER to vacuoles via Golgi (Parker 1980, Parker 1982, Loussert et al. 2008) or they may accumulate within ER then move to protein bodies (Campbell et al. 1981, Parker 1982). A coalescence of protein bodies is usually observed during



Shewry and Halford 2003

Figure. 12. Chromosomal localization of genes encoding HMW and LMW-GS.

protein accumulation. The final protein bodies (1-2 microns of diameter) are disrupted upon grain dehydration giving a protein matrix surrounding the starch granules.

In wheat, a network of regulatory proteins known as TF-like (transcription factor) storage protein activator (SPA) and prolamin-box binding factor (PBF) are involved to control prolamin gene expression in a specific temporal and spatial pattern (She et al. 2011). Alpha/beta gliadins and LMW glutenin-encoding genes are only expressed during seed development stages, two expression peaks were discerned during protein accumulation one at10 DPA and the second at 20 DPA (Kawaura et al. 2005). HMW-GS were reported to peak at 12 DPA and expressed preferentially in the lobes of starchy endosperm (Lamacchia et al. 2001).
At transcript level, LMW-GS initial synthesis can progress with that of HMW-GS (Johnson et al. 2006, Tosi et al. 2009). Other reports suggested that minor amounts of LMW-GS were synthesized after HMW-GS at about 14 DPA (Loussert et al. 2008). Takemoto et al. 2002 indicated that in rice, cystein rich prolamins were synthesized throughout development while cystein poor prolamins appeared abundantly at later stages in the endosperm.

3.3 Impact of Wheat proteins on its Quality

Wheat quality parameters are at least as wide as their uses. Wheat that is suitable for one particular use may possess certain qualities that render it entirely unsatisfactory for another use. The various kernel parts also have different uses; for example, bran is used for feed, germ as a diet supplement and the endosperm for flour. Quality criteria in most cases refer to the characteristics of the endosperm (Pomeranz 1988).
On the basis of botanical characterization wheat species and varieties are the main determinants of quality. Weight per unit volume, kernel weight, its size, shape, hardness, vitreousness, impurities and kernel condition (damaged or not), color and milling quality are the physical parameters for quality determination. Chemical characteristics which have impact on quality include moisture content, protein content and quality, alpha amylase activity, fat acidity, crude fiber and ash.
Protein content of wheat ranges from 6 to 20% depending upon variety and environmental conditions during growth. Proteins are useful markers of quality (Payne 1987). Several studies

Gene	Allele(s)	Main phenotype	Reference
HMW-GS			
Dx, Dy	Various	Dough strength	[Marchylo et al., 1989a] and [Marchylo et al., 1989b], [D'Ovidio etal.,1994] and [De Bustos et al., 2001]
Dx5	*Glu-D1d*	Dough strength	[Smith et al., 1994], [D'Ovidio and Anderson, 1994], [Ahmad, 2000],[Ma et al., 2003] and [Radovanovic and Cloutier, 2003]
Ax, Ay	Various	Dough strength	Lafiandra etal. (1997)
Ax null	*Glu-A1c*	Dough strength	[D'Ovidio et al., 1995] and [Lafiandra et al., 1997]
Ax	Various	Dough strength	De Bustos et al. (2000))
Dy10/Dy12	*Glu-D1d/a*	Dough strength	Ahmad (2000)
Bx7/Bx17	*Glu-B1b*	Dough strength	[Ahmad, 2000] and [Ma etal., 2003]
Ax2B	-	Dough strength	Juhasz etal. (2001), using marker of D'Ovidio and Anderson (1994)
Ax1, Ax2	*Glu-A1a,b*	Dough strength	[Ma etal., 2003] and [Radovanovic and Cloutier, 2003]
Bx7,Bx7	*B1al*	Dough strength	[Radovanovic and Cloutier, 2003] and [Butow et al., 2003]
Bx	*B1 al, b, i*	Dough strength	Butow etal. (2004)
LMWGS			
	Glu-B1	Pasta quality	[D'Ovidio, 1993] and [D'Ovidio and Porceddu, 1996]
i-type LMWGS	*Glu-A3 a–f*	Dough strength and extensibility	Zhang et al. (2004)
	Glu-A3	Dough strength and extensibility	Devos et al. (1995)
Puroindolines			
pinA	*PinA-D1a,b*	Grain hardness	Gautier etal. (1994)
pinB	*PinB-D1a,b*	Grain hardness	Giroux and Morris (1997)
pinB	*PinB-D1c*	Grain hardness	Limello and Morris (2000)
pinB	Various	Grain hardness	Morris (2002)
Unknown		Grain hardness	Partridge etal. (2002)a
GBSS1			
GBSS1-4A	*Wx-B1a*	FSV	[Briney etal., 1998] and [Gale et al., 2001][a]
GBSS1-7D	*Wx-D1a*	FSV	Shariflou etal. (2001)
GBSS1	Various	FSV	McLauchlan etal. (2001) and Nakamura etal. (2002)
GBSS-7D	*Wx-D1e*	FSV	Yanagisawa etal. (2003)
Wheat/rye translocations			
Rye chromatin		Dough properties	[Francis et al., 1995] and [Koebner, 1995]
ω-secalin, LMWGS	1B/1R	Dough stickiness	[de Froidmont, 1998], [Zhang et al., 2003] and [Andrews et al., 1996]
ω-secalin	*Sec-2*(Rye 2RS)	Dough stickiness	Skerritt et al. (1996)

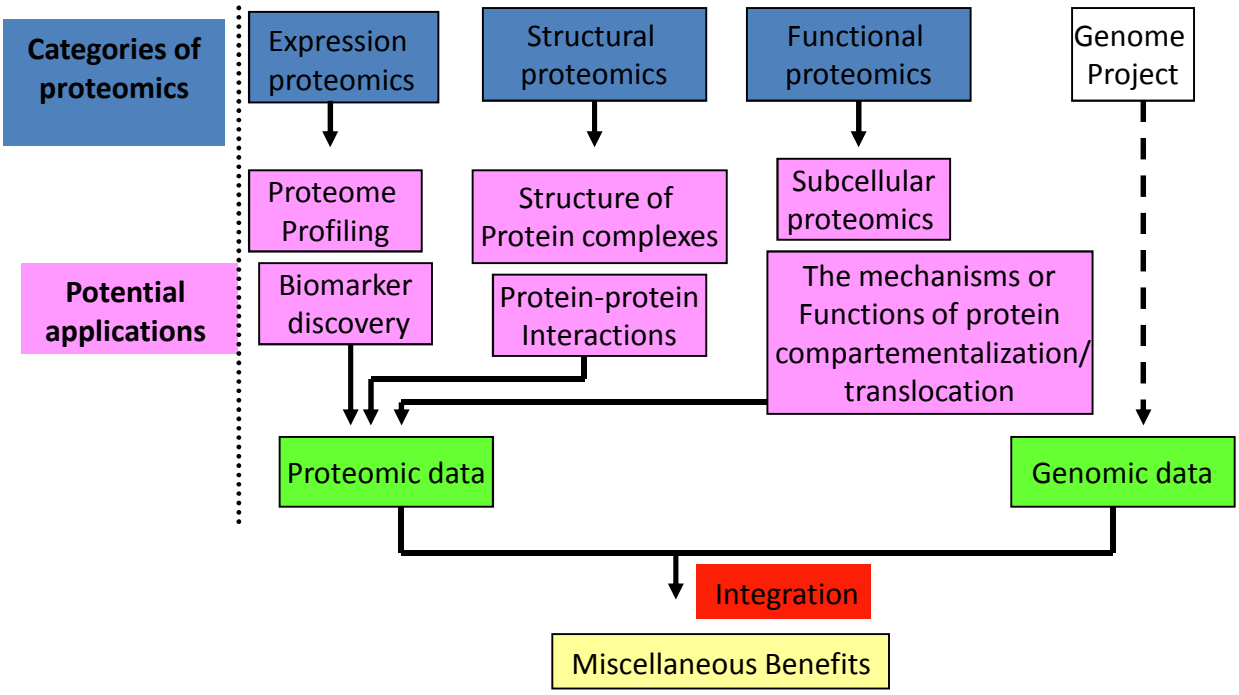
Gale, K.R., 2005

Table 9: List of diagnostic molecular markers for quality traits in wheat.

indicate that in bread making quality, 20-60% of the variability between varieties can be accounted, by variation in grain storage proteins (Payne 1987, Lukow et al. 1989, Kolster et al. 1991, Worland, and Snape 2001). Gluten quality is largely a varietal characteristic, in addition to environmental factors. Among glutenin polymers HMW-GS are important components that play a major role in determination of the unique visco-elastic properties of wheat doughs (Halford et al. 1992, Nieto-Taladriz et al. 1994). There are three loci for HMW-GS: Glu-A1, Glu-B1, Glu-D1 (Figure 12). Each locus encodes an x- and y- type HMW-GS, where x- type encodes HMW-GS of higher molecular weight than y-type. Of the three loci, Glu-D1 has the most significant effect on dough strength. Extra cystein residues present in N-terminal unique domain of 1Dx5, result in stronger dough (in combination with 1Dy10) (Lafiandra et al. 1993, Chen et al. 2012). Recently the sequence named 1Dx3 in *Aegilops tauschii*, was found to have similarity with 1Dx5, suggesting that it might have positive effects on dough quality (Wang et al. 2012).

The effect of allelic variation of LMW-GS on wheat quality attributes has received far less attention than that of HMW-GS, as scoring of these sub-units on SDS-PAGE is relatively difficult. LMW-GS are thought to play a major role in determining extensibility and to lesser extent the dough strength, both directly and via their interaction with HMW-GS (Gupta and Macritchi 1994, Luo et al. 2001). Groos et al. 2004, reported QTL's for dough strength on chromosomes 3A and 7D, suggesting that non-glutenin parts also have some influence on dough quality.

Bread making processes are dependent upon an optimal mixture of gliadins and glutenins (gluten), and disulphide bonds are the key factor in determining gluten properties (Shewry and Tatham 1997). They link two cysteine residues, either by intrachain (within a single protein) or by interchain (between proteins) bonds. Gliadins are mostly monomeric and contain either no cysteine (ω5- and ω1,2-gliadins) or disulfide bonds within a single protein. α-gliadins contain six and γ-gliadins, six to eight, cysteine residues located in the C-terminal domains. They form three and four homologous disulfide bonds between proteins (Müller and Wieser 1995, Lutz et al. 2012). A comprehensive review of wheat quality determining factors was provided by Gale K.R. and is presented in Table 9, which also includes some additional markers to that are discussed above in this section (Gale 2005).



Modified from LAU , et al. 2003

Figure 13: Categories and potential applications of proteomics.

4 Proteomics to Study Expression of Candidate Genes

4.1 Proteomics

The proteome is the set of expressed proteins in a given type of cells or an organism at a given time under defined conditions (Wilkins et al. 1996). In technical term the study of proteome called proteomics, can be classified into three groups according to Lau et al. 2003 (Figure 13).

i) Expression or Differential proteomics. A quantitative study of protein expressions between samples is studied under this class. Using this approach protein expression of the entire proteome (ideally) or subsets of proteome can be compared. In addition novel proteins (e.g. disease-specific biomarkers) can also be identified.

ii) Structural proteomics. Under this group we map the structure of proteins or protein complexes present in a sub cellular localization or an organelle (Blackstock and Weir 1999).

iii) Functional proteomics. Relative quantification of proteins, targeted by their function is studied in this class. Proteins relations to each other and to the specific biological condition of the cell also come under its area (Monti et al. 2005).

Analysis of protein profiles at sub cellular sites is helpful in understanding the functional organization of cells at the molecular level. The combination of protein identification by mass spectrometry with fractionation techniques for the enrichment of particular sub cellular structures is termed as 'sub cellular proteomics' (Dreger 2003).

4.2 Quantitative Proteomics

In comparative proteomics, quantitation is an intrinsic requirement. Most quantitative proteomics methods are common among prokaryotes and eukaryotes such as humans, animals, yeast and plants as well. In plant quantitative proteomics, low protein concentration, complex and large plant genomes, are some frequent challenges. In contrast, control of growth conditions and easy establishment of cell culture are advantageous compared with animals (Bindschedler and Cramer 2011). Classification and different methods used in quantitative proteomics are presented in

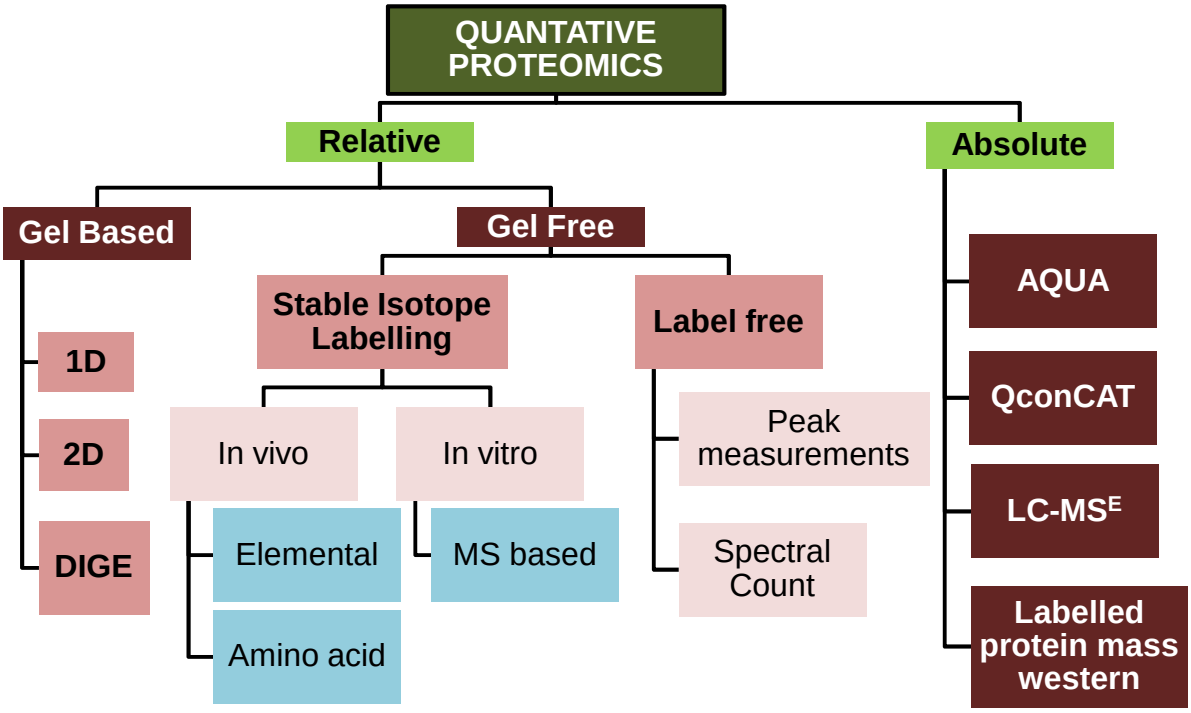


Figure 14: Classification of quantitative proteomics.

(Figure 14). Relative and absolute quantifications can be used according to analysis and required results.

4.2.1 Relative quantification

Relative quantification is well established, using different strategies comparative levels of proteins or individual peptides can be measured in two or more samples.

4.2.1.1 *Gel based relative quantification*

Gel based Techniques need gel preparation to visualize the protein spot, before their identification. 1D (one dimensional) is mostly to compare different bands, or can be coupled to 2DE (two dimensional electrophoresis) for specific band visualization on 2D plan. DIGE (difference gel electrophoresis) is modified form of 2DE and was first described by Ünlü et al. 1997. The difference from 2DE is that protein samples (up to 3) are labeled with fluorescent dye before 2DE, they are mixed and a single gel is formed from this mixture.

4.2.1.2 *Gel free relative quantification*

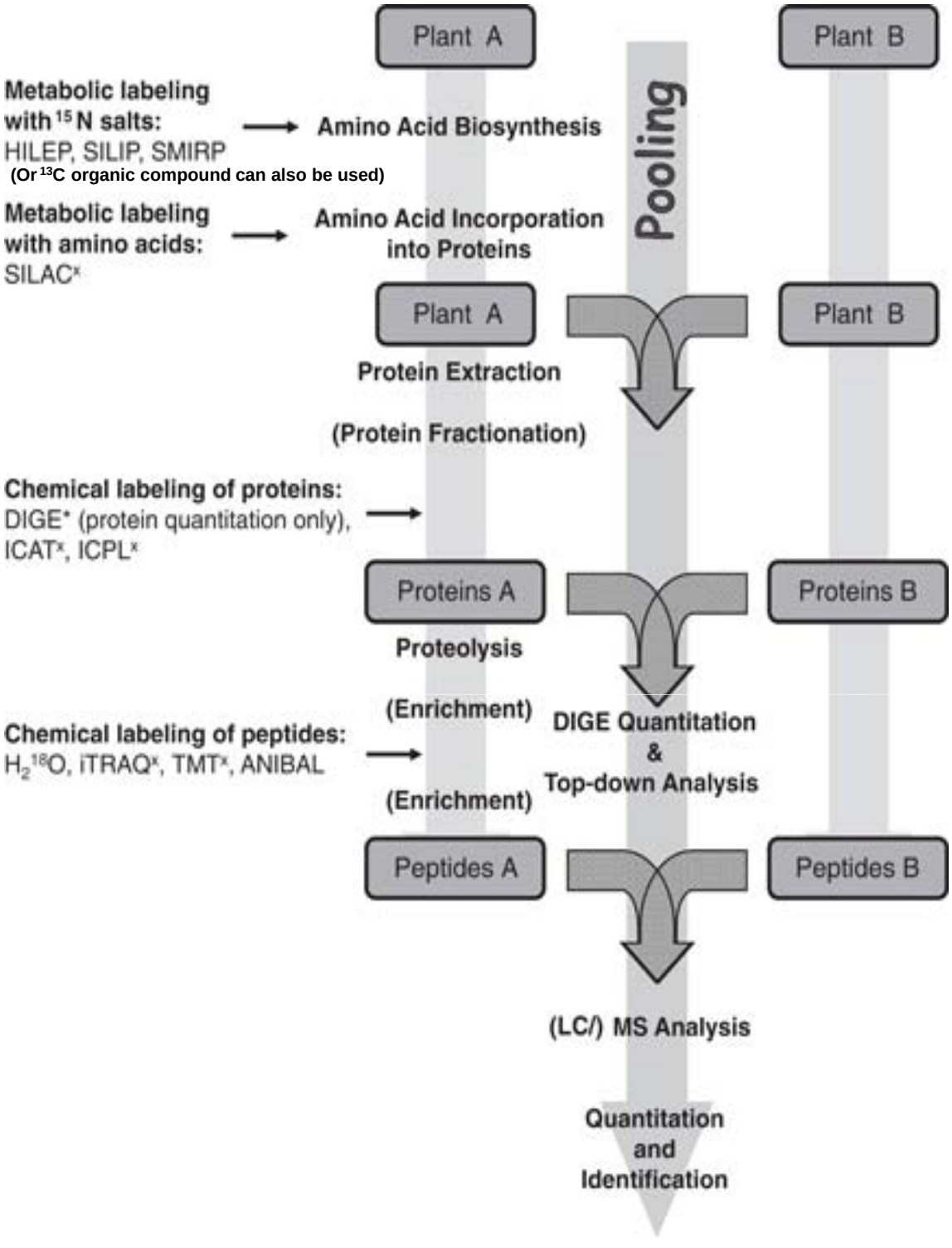
Gel free techniques permit quantification by using stable isotope labeling or label free methods. Classification and different strategies used under labeled and label free category are shown in Figures 14 and 15.

4.2.1.2.1 Gel free relative quantification using labeling methods

Commonly used labeling methods described are:

Metabolic Labeling (*in vivo*):

Elemental	HILEP Hydroponic Isotope Labeling of Entire Plants (Bindschedler et al. 2008)
	SILIP Stable isotope labeling in planta (Schaff et al. 2008)
	SMIRP subtle modification of isotope ratio proteomics (Whitelegge et al. 2004)
Amino acids	SILAC stable isotope labeling with amino acids in cell culture



Modified from Bindschedler, L.V., et al. 2011

Figure 15: Stable isotope labeling methods and their sample processing workflow.
* Indicates DIGE quantitation is independent of MS analysis.
^{x} indicates the labelling techniques that allow multiplex analysis (in the same experimental run, more than two samples can be compared

In elemental quantification ^{15}N Salts or ^{13}C organic compounds are used for labeling. This method is used successfully in mammals (Gustavsson et al. 2005) *E.coli* (Ross 2004), and in plants (Engelsberger et al. 2006, Palmblad et al. 2008). In some cases whole higher organisms can be labeled (Krijgsveld et al. 2003). For labeling of amino acids, lysine, arginine or leucine are generally used (Gruhler et al. 2005, Mann 2006, Schütz et al. 2011).

Chemical labeling (*in vitro*):

Of Proteins
- DIGE Difference gel electrophoresis (Ünlü et al. 1997)
- ICAT isotope coded affinity tag
- ICPL isotope-coded protein labeling (Lottspeich and Kellermann 2011)

Of Peptides
- $H_2{}^{18}O$ DNA is labeled with $H_2{}^{18}O$ (Schwartz 2007)
- iTRAQ isobaric tag for relative and absolute quantification (Pichler et al. 2011)
- TMT Tandem mass tag (Pichler et al. 2011)
- ANIBAL aniline and benzoic acid labeling (Panchaud et al. 2008)

4.2.1.2.2 Label free quantification

Peak intensity measurements and spectral counts (for relative and absolute quantitation) are the two commonly used methods in this class of quantification. To overcome possible errors because of run-to-run variations, it is essential to use the methods with good mass accuracy, highly reproducible nano-HPLC separation and very delicate and sensitive computational tools (Domon and Aebersold 2006, Koubaa et al. 2012).

4.2.2 Absolute Quantification

Absolute quantification is less well established compared to relative, and is performed using internal standards (usually peptide surrogates). In this method protein abundances are ranked in

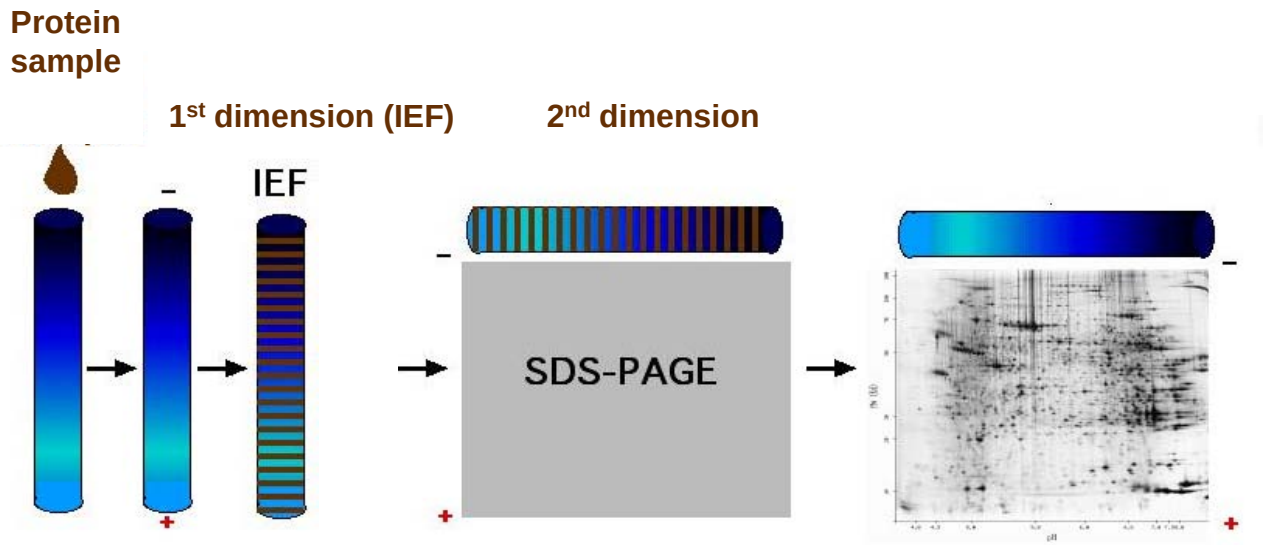
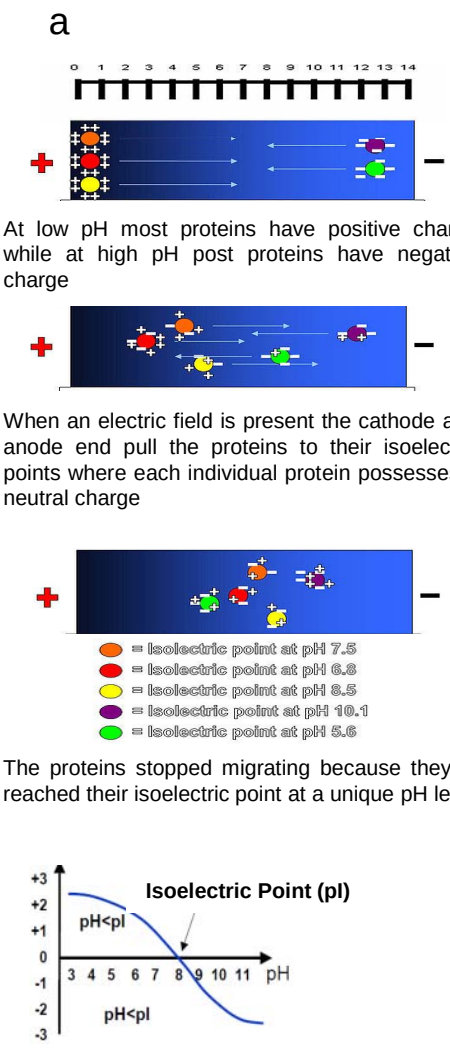


Figure 16: Principle of 2-dimensional electrophoresis. Iso electro focusing (1st dimension), Mass dependent separation of proteins (2nd dimension).



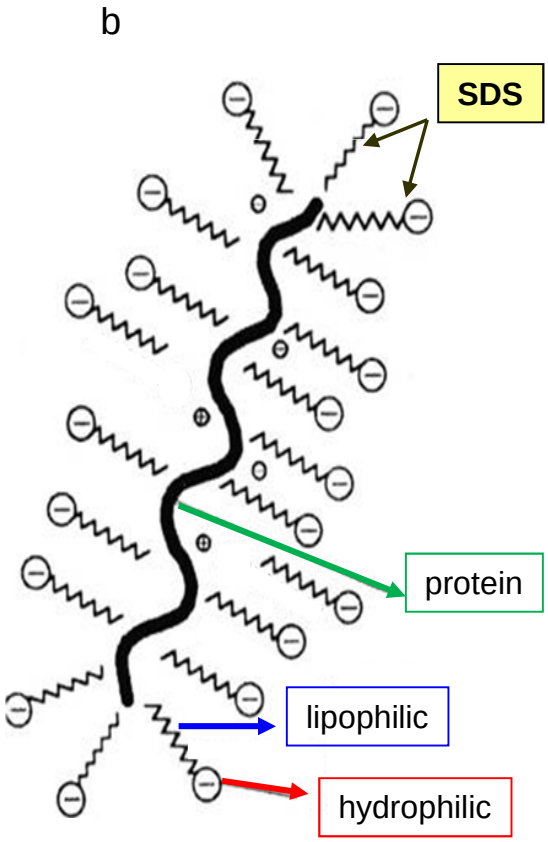


Figure 17: a) IEF work flow
b) SDS molecule and protein interaction

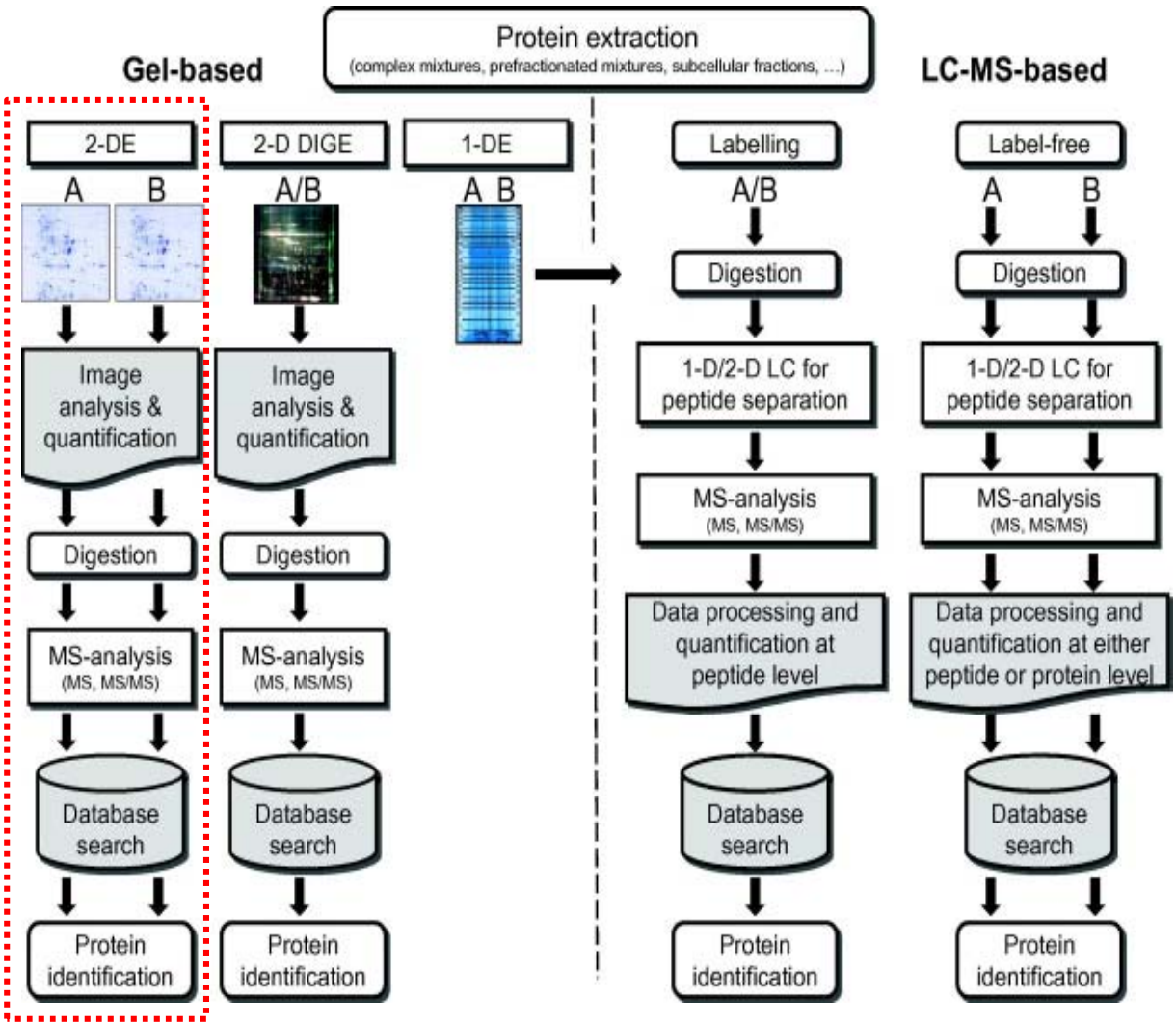
order and assessment of stoichiometry can also be performed. This approach facilitates targeted proteome analysis (Peltier et al. 2006, Huillet et al. 2012).

2D-PAGE (2 Dimensional-Poly-Acrylamide Gel Electrophoresis)

Two-dimensional Electrophoresis (2-DE) is separation of proteins using Immobilized pH Gradient (IPG), for their separation firstly according to the iso-electric point and secondly according to their molecular weight by the use of polyacrylamid gel (Görg et al. 1988, Gorg et al. 2000). 2DE is combined with protein identification using MS (mass spectrometry) (Aebersold and Mann 2003). Inspite of advancement in proteomic technologies, this technique is still a powerful tool for protein identification (Rabilloud 2002, Görg et al. 2004).

In 2DE, proteins are separated and identified in two dimensions oriented at right angles to each other (Figure 16). In first dimension IEF (Isoelectric focusing) separates proteins on the basis of their net charge. Proteins are amphoteric molecules; they carry either positive, negative, or zero net charge, depending on the pH of their surroundings. The net charge of a protein is the sum of all the negative and positive charges of its amino acid side chains and amino and carboxyl-termini. The isoelectric point (pI) is the specific pH at which the net charge of the protein is zero. If net charge is plotted against the pH of its environment, the resulting curve intersects the x-axis at the Isoelectric point (Figure 17 a).

In second dimension further separation of proteins by their mass is obtained using Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) (Figure 17 b). SDS ($C_{12}H_{25}NaO_4S$) is a strong detergent agent used to denature native proteins to unfolded, individual polypeptides. The lipophilic part of SDS attaches to the lipophilic amino acids in the protein, while hydrophilic part (negative charge) of SDS shows to the outside of the protein. Consequently all proteins have the same stretched rod like shape and intrinsic charges of polypeptides become negligible when compared to the negative charges contributed by SDS. These negatively charged protein complexes migrate to the anode and their separation is only mass dependent i:e small proteins migrate faster than large proteins.



Matros, A., et al. 2011

Figure 18: Schematic workflow presentation of the different approaches used in quantitative proteomic studies.

2D-PAGE without using SDS is called Native PAGE, where proteins are not denatured before electrophoresis. For most cases, protein complexes remain associated and folded as they would be in the cell, this may result in gels where separation of proteins is not clear. Different steps involved from sample preparation to protein identification for 2D along with other gel based and LC-MS based techniques are summarized in Figure 18 (Matros et al. 2011).

4.3 Mass Spectrometry

Mass spectrometry is a very important analytical method used in practically all chemistry laboratories the world over. Previously, only fairly small molecules could be identified, but John B. FENN and Koichi TANAKA have developed methods that make it possible to analyse biological macromolecules as well. They were awarded the Nobel Prize of chemistry in 2002 concerning this effort. There are three fundamental parts in a mass spectrometer (MS);
i) the ionization source converts neutral molecules to ions by adding or taking away one or more protons. This result in single or multiple charges of ions which are easier to control and detect than neutral molecules. ii) the analyzer separates ions according to their mass to charge (m/z) ratio under an electromagnetic field. iii) the detector measures the relative abundance of each ion present. Finally, signals are sent to the data system and are formatted in m/z spectrum. Many different types of ion source, analyzer and detector exist, depending upon the type of analysis to be performed. Here, I will shortly describe only MALDI-TOF and LC-MSMS, which were used in our experiments.

4.3.1 MALDI-TOF (Matrix-Assisted Laser Desorption -Time of flight)

MALDI ion source is coupled with TOF type mass analyzer for this type of mass spectrometry analysis. MALDI is a soft ionization technique using a laser beam and involving two steps. In the first step, matrix molecules are ionized by laser flash, in the second step, sample molecules are ionized by proton transfer from matrix.
Time-of-flight (TOF) is the least complex mass analyzer in terms of its theory. Ions are given a defined kinetic energy and allowed to drift through a field-free region (0.5 to several meters). The

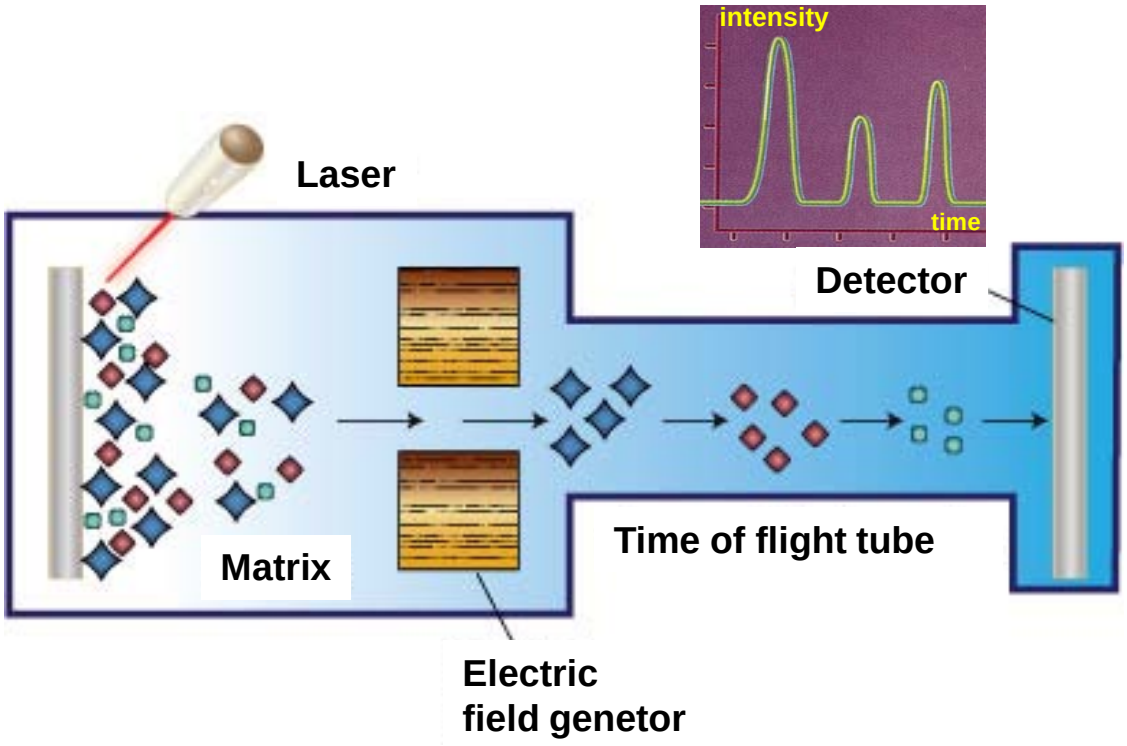


Figure 19: Simplified schematic of MALDI-TOF mass spectrometry.

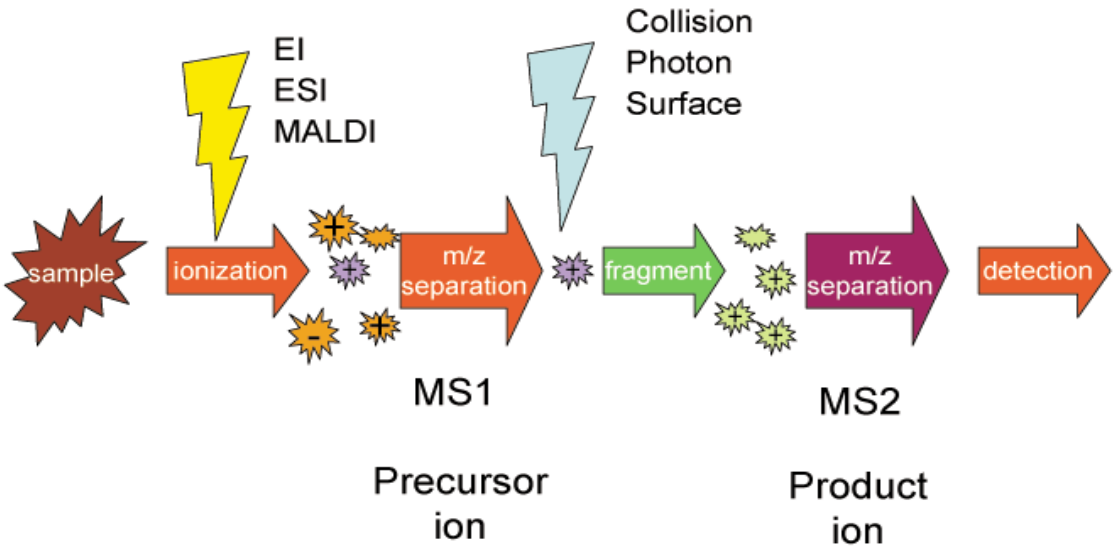


Figure 20: Schematic of Tandem mass spectrometry.

time ions take to arrive at the detector is measured and related to the m/z ratio. Lighter ions will reach the detector first (Figure 19).

4.3.2 LC-MS (Liquid chromatography mass spectrometry)

Liquid chromatography mass spectrometry (LC/MS or LC-MS) separates compounds chromatographically before they are introduced into the ion source and mass spectrometer. The mobile phase is liquid, usually a mixture of water and organic solvents.

4.3.3 Tandem mass spectrometry (MSMS or MS^2)

Coupling mass spectrometers in series provides a new technique named Tandem mass spectrometry, MSMS or MS^2 (McLafferty 1981). In the first mass spectrometer, the targeted compound is selectively ionized, and its characteristic ions are separated from most others of the mixture. The selected primary ions are then decomposed by collision. The final mass analyzer, from the resulting products, selects secondary ions characteristic of the targeted compound (Figure 20). In the case of proteins, the masses of the identified peptides are measured in first MS. These selected peptides are then passed for further fragmentation in a second mass analyzer where the fragment ion masses are measured. The amino acid sequence can be inferred from mass differences between peaks, as fragmentation occurs most often along the peptide backbone.

5 Transcriptomics to Study Expression of Candidate Gene

Transcriptomics is a term given to the set of all transcripts or messenger RNA (mRNA) molecules produced in cells. It can also be applied to the specific subset of transcripts present in a particular cell or the total set of transcripts in a given organism. It is not only used to capture the spatial and temporal gene expression within the tissues, but is also commonly used to increase our understanding about molecular events which control biological processes.

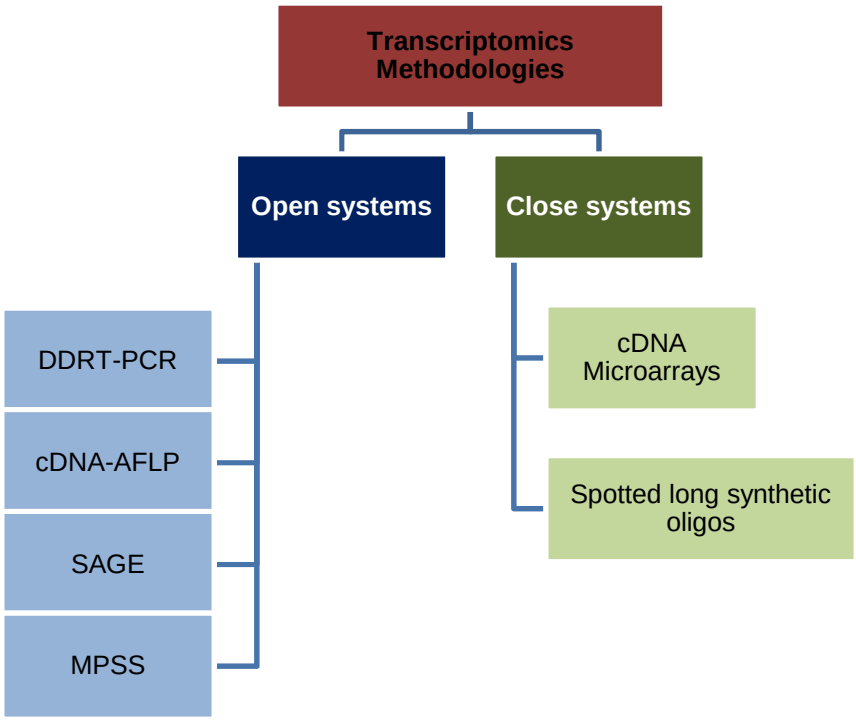


Figure 21: General methodologies used in transcriptomics.

5.1 Methods to Study Transcriptional Expression

A range of methodologies are available in plants for analysis of gene expression, that fall broadly in two categories, i.e Open and Close systems (Figure 21) (Leader 2005).

5.1.1 OPEN SYSTEMS

In these systems, prior cloning or sequencing of genes is not needed and, in principle, can generate novel expression information about any transcript in an organism. Briefly in plant biology open systems include:

5.1.1.1 *RNA finger printing techniques (DDRT-PCR and cDNA-AFLP)*

Differential display reverse transcriptase polymerase chain reaction (DDRT-PCR) was one of the first methods for analysis of differential gene expression on the scale of the full transcriptome (Liang and Pardee 1992). To identify the differentially-regulated genes, DNA fragments which show differential expression between conditions, are subsequently cloned.

cDNA-AFLP, amplified fragment-length polymorphism, is also a PCR based tool. It is particularly useful technique for detecting rare, differentially-regulated transcripts which are often missing from microarray datasets or from RNA fingerprint experiments. Volkmuth et al. developed a high throughput protocol for cDNA-AFLP using Arabidopsis model plant to extend the usefulness of the technique as a true genomic tool (Volkmuth et al. 2003).

5.1.1.2 *EST, Expressed Sequenced Tags*

With the emergence of new open-system strategies for cDNA analysis, it became possible to rapidly obtain the fragments of sequences from hundreds of randomly selected cDNA clones (Adams et al. 1991). Because cDNA represents the mRNA set, these short sequences were dubbed expressed sequence tags (ESTs).

5.1.1.3 *SAGE, serial analysis of gene expression*

SAGE is a conceptually related but technologically distinct and more systematic EST-like approach. It is based on generating short transcript tags (13-34 mers) by restriction of cDNA molecules and ligation to larger fragments prior to sequencing. After sequencing, the abundance of a specific tag is compared with the total number of sequenced fragments, reflecting the relative expression level of its corresponding gene transcript (Velculescu et al. 1995).

5.1.1.4 *MPSS, massively parallel signature sequencing*

MPSS generates similar data to SAGE but at much higher throughput since up to one million individual tags can be sequenced simultaneously in a single experiment. The dynamic range for MPSS is typically from below 10 transcripts per million (tpm) to more than 50,000 tpm.

5.1.2 CLOSED SYSTEMS

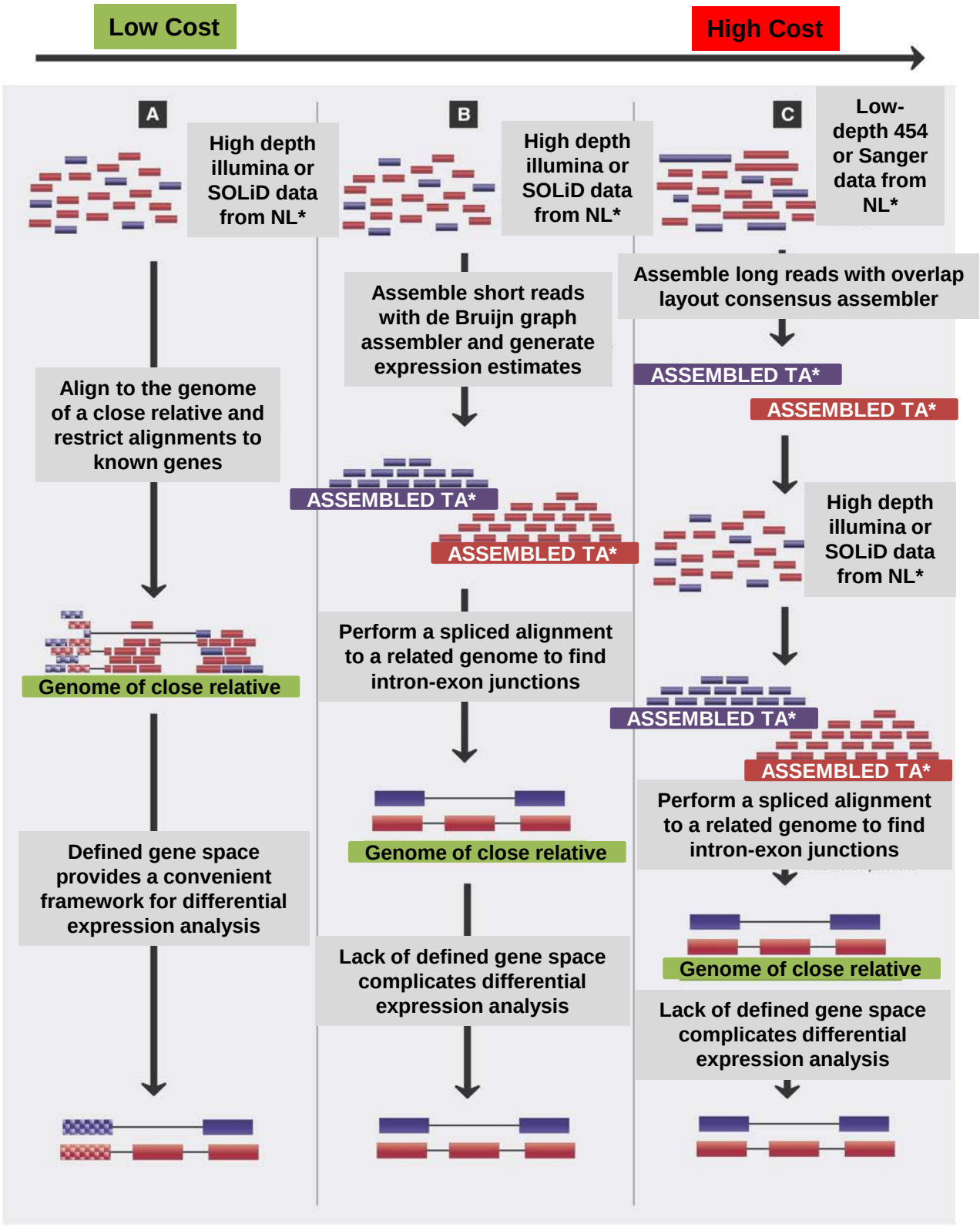
For these analyses a prior sequence is required and includes:

-cDNAs, in situ synthesized oligo probes

-Spotted long synthetic oligos

With accumulation of thousands of cDNA and EST sequences, automation and miniaturization technologies were developed to place these sequences in closely packed arrays that could be used as cDNA hybridization targets so as to allow the expression of thousands of genes to be tracked simultaneously (Lockhart et al. 1996). Importantly, cDNA arrays do not require a full genome sequence and they can be generated relatively easily from pre-sequenced or anonymous cDNA clones (Zhao and Bruce 2003).

Due to limitations of both closed and open systems, the new promising technique available is the Roche 454 sequencing-by-synthesis (Roche Diagnostics Corporation, Basel, Switzerland), which allows the potential for deep transcriptome coverage and produces long reads that are more amenable to annotation and specific genome assignment in the hexaploid genome (Coram et al. 2008).



Ward et al , 2012

Figure 22: Schematic comparison of three possible options for the analysis of the transcriptome data in non model organisms. (**A**) Align-then-assemble; **(B**) Assemble-then-align variant 1; (**C**) Assemble-then-align variant 2.
NL*=non normalized libraries, TA*=Transcript

5.2 Wheat Transcriptomics Studies

In recent years, transcriptomics has been used in understanding the molecular basis of wheat grain, its development and underlying mechanisms. Starting from the vegetative phase, map-based cloning of vernalization genes has been achieved by genetic analysis (Yan et al. 2003, Yan et al. 2004, Yan et al. 2006). For spring growth habit in tetraploid wheat, a novel retrotransposon was identified using qPCR based techniques (Chu et al. 2011). Transcription profiling of winter and spring wheat cultivars have been performed by Winfield et al, 2009.

During grain development larg scale EST sequencing (3500 ESTs) was first described by Ogihara et al. 2003 and was extended to include an additional 16, 000 ESTs by Wilson et al. 2004. cDNA arrays from six stages of grain development containing up to 7800 sequences are also available (Laudencia-Chingcuanco et al. 2006, Laudencia-Chingcuanco et al. 2007). Wheat gene Chip allows simultaneous investigation of 55,000 gene transcripts, and was used by Wan et al. 2008, to study grain development profiles. Early stages of grain development were studied by Drea et al. 2005, while McIntosh et al. 2007 studied wheat caryopsis using SAGE at 5 stages of development, spanning 8 to 40 DPA.

Wheat has a huge non-sequenced genome, but sequence information about model plants (such as brachypodium, arabidopsis, rice and sorghum) can be transferred to non-sequenced genomes for agricultural improvement (Spannagl et al. 2011). To this end comparative transcriptomic analysis of gene expression evolution among three Poaceae members revealed that highly and broadly expressed genes tend to be conserved at the coding sequence level while narrowly expressed genes show accelerated rates of sequence evolution. This data also revealed higher correlation among expression patterns of syntenic orthologs as compared to non-syntenic ones (Davidson et al. 2012). Ward et al described the three different transcriptomics strategies for non-model plant species (Ward et al. 2012) (Figure 22).

(A) Align-then-assemble: Short-read data can be directly aligned to the genome of a close relative, using predefined gene models from the organism's annotation. This is particularly useful for analysis of experiments because comparisons between gene models are unambiguous. However, exons and whole genes can be missed when reads fail to align (checkered exons).

Table 10: Some studies on grain development of cereals and non-cereals. Only proteomics or transcriptomics studies are given.

Reference	Plant Material Used	Variety	Technique/ Method of study used
WHEAT			
Nadaud I. 2010 Proteomics, 10:2901-10	0-280°Cd (11 stages)	*T.aestivum* Récital	SDS-PAGE ,2DE, MALDI-TOF-MS
Shewry P. R. 2009 J. Cereal Sci., 50: 106-112	10-42DPA (8 stages) and Mature Grain	*T.aestivum* Hereward	SDS-PAGE (1D), Transcriptomics and HPLC
Wan Y 2008 BMC Genomics. 9 :121	6-42DAA (10 stages)	T.aestivum Hereward	Transcriptomics
RICE			
Sheng Bao Xu 2008 *Plant Physiology* 148:908-925	2-20 DAF grains (10 stages)	*Oryza sativa* Nipponbare	2DE –MALDI Tof
BARLEY			
Finnie C. 2002 Plant Physiology, 129:1308-1319	Grain filling and seed maturation	*Hordeum vulgare*	2DE –MALDI Tof
BRACHYPODIUM			
Opanowicz M 2011 Journal of Experimental Botany 62:735-748.	0-20DAA (9 stages)	*Brachypodium distachyon*	Cytological studies and SDS-PAGE
MAIZE			
Mechin V., 2007 Plant physiology, 143:1203-1219	4-40DAP (7 stages)	*Zea mays* (inbred line F2)	2DE, LC-MS/MS
NON-CEREALS			
Faiza Tebbji 2010 *BMC Plant Biology* 10:174	Ovules at 12 stages 0-22 DAP	*Solanum chacoense*	cDNA microarray
Agrawal G.K. etal., 2008 *Plant Physiology* 148:504-518	2-6 WAF (5stages)	*Glycine max and Brassica napus*	2DE and MudPIT with nESI-LC-MS/MS
Hajduch M, etal., 2006 Plant Physiol,141: 32–46	2-6 WAF (5stages)	*Brassica napus*	2DE-MALDI Tof and LC-MS
Norma L. H. 2009 *Plant Physiology* 151:857-868	2-6 WAF (5 stages)	*Ricinus communis*	2DE-LC-MS
Gallardo K., 2003 Plant Physiol, 133: 664–682	8-44 DAP (10 stages)	*Medicago truncatula* cv Jemalong	2DE-MALDI Tof and ESI-MS/MS
Shi Y 2010 Journal of Experimental Botany 61 : 2367-2381	6 stages (every 2 weeks)	*Cunninghamia lanceolata*	2D-DIGE-MALDI Tof
Dam S 2009 *Plant Physiology* 149:1325-1340	7-43 DAF (10 stages)	*Lotus japonicus*	DNA sequencing, 1 and 2DE, MALDI-MS and LC-MS
Alonso AP 2007 The Plant Journal, 52: 296-308	12 -17 DAP (5stages)	*Helianthus annuus* L.	GC-MS

(B) Assemble-then-align variant 1: Short-read data are assembled into transcripts in a manner that allows integrated detection of gene expression, but lack of defined gene space can result in ambiguous comparisons in the experimental context. The genome of a close relative can help define comparisons among large numbers of transcripts. The approach also requires additional sequencing depth compared to the align-then-assemble approach.

(C) Assemble-then-align variant 2: Longer 454 reads or Sanger reads are assembled into transcripts. 454 data sets are typically normalized prior to sequencing the initial assembly; therefore, expression information is lost. Again, lack of defined gene space can result in ambiguous comparisons in the experimental context, but the genome of a close relative can help to define comparisons among large numbers of transcripts. The approach is the most expensive in terms of sequencing costs and computational time.

6 Integration of Proteomics and Transcriptomics data

In plants, considerable advancement in "omics" techniques have been made in recent years for study of plant phenotype diversity. For study of changes in genome, microarray and serial analysis of gene expression (SAGE), approaches have provided important advances in many plant species including *T. aestivum*, during the past decade (Table 10). These profiling technologies do not require selection of the important genes in advance and provide an unbiased approach to identify genes whose regulation is changed during environmental perturbation of the cell, tissue or organism (Waters et al. 2006). Proteomics plays a central role in systems biology as it complements transcriptome and metabolome data. Many studies on grain have been performed to describe proteome profiling during development both in cereals and non-cereals (Table 10).

The steady-state level of a protein presents in a given cell or tissue is thought to be modulated by post-transcriptional mechanisms, including translation, post-translational modifications. Data of gene expression patterns are therefore insufficient for the prediction of protein abundance. To gain insight into a broad range of biological processes and to understand the organization principles of cellular functions, integration of complementary information contained in proteomic and mRNA data sets is needed. Although this comparison poses several important challenges, these efforts should provide considerable advancement concerning fundamental mechanisms underlying physiology, development, and the emergence of disease. The major challenge is

inadequacy of available statistical tools to compensate for biases in data collection methodologies which result in poor correlations between data sets. This type of proteome and transcriptome correlation study is well described in sequenced organisms particularly non-plants and is less well developed in those that have not yet been sequenced. Some recent examples are given here:
In aphids this combined proteomic and transcriptomic approach was used to study photoperiodic regulation (Le Trionnaire et al. 2009). Zhang et al published a comprehensive review which covered the "interactomics" studies made on microbial organisms (Zhang et al. 2010). A study on proteome and transcriptome variation in mouse resulted in an average correlation of 0.27. This varied depending on the cellular location and biological function of the gene (Ghazalpour et al. 2011). In Zebra fish, natural routes of an infectious disease have been described, also using proteome and transcript profiles in control and infectious individuals (Encinas et al. 2010). Similarly the same approach was used in flatfish (*Senegalese sole*) at various stages of testes, throughout spermatogenesis to improve understanding of the molecular mechanisms responsible for the successful production of functional spermatozoa (Forne et al. 2011).
In arabidopsis, General Linear Modeling was used to evaluate the protein/transcript expression patterns at five stages of development. 56% of expression pairs showed a significant correlation using this statistical assessment (Hajduch et al. 2010). Genomic and proteomic approaches were applied to identify candidate genes and proteins involved in antimicrobial defense and (heavy) metal uptake and translocation in lupin (*Lupinus albus* L.) roots at 5, 10, 15 and 20 days post-emergence (Tian et al. 2009). In medicago (*Medicago truncatula*) 50% protein/transcript were found correlated at six stages from 12-36 days after pollination (Gallardo et al. 2007).

Advances in genome sequencing using modern technologies would facilitate this type of comparison, and would result in more reliable correlations among "omics" data.

THESIS OBJECTIVES

The present study was based on the following two main themes:

1: Grain tissue proteomics

Many studies have been published to describe the physiology of wheat grain development. Proteomic studies of wheat grain development have also been described, but only divided into a few stages which span the development with large temporal durations. The research reported here concerned the development of two important grain tissues (endosperm and peripheral layers) over very short temporal durations, with twenty one stages from 0 to 1006°Cd for endosperm and at fifteen stages from 0 to 700°Cd for peripheral layers.

Metabolic proteins of starchy endosperm have already been studied but only at two stages of development, and a proteome map of peripheral layers is described for the first time by our study. The main aims of our study were following:

- What are the principle phases, during development of peripheral layers and endosperm?
- How many profiles could result during development of both tissues?
- What are the main functional classes for differentially expressed identified proteins?

These classes were analysed i) to understand how these functions vary among profiles. ii) to develop a proteome map for each profile and to study what are the variation (differences/ similarities) among profiles. In the proteomic study of starchy endosperm proteins, we investigated some additional questions:

What are the different localizations which can be predicted for proteins identified in starchy endosperm?

How does the abundance of stress/ defense category proteins change during development? What sub-categories exist in the stress/ defense class and how do they vary?

This data from these two proteomic studies enabled us to compare proteomic changes which occur in the two tissues at a given stage: comparison of expression profiles, functional profiles and total proteins identified. This additional information is provided in “Chapter 5”.

2: Integration of Proteome and Transcriptome data

In recent years, research has been directed towards integration of different datasets for better understanding of fundamental mechanisms underlying physiology, development, and the emergence of disease. In plants, regarding proteome and transcriptome data linkage, there have been very limited studies compared to other organisms. We proposed to combine proteome data (obtained by our study) and transcriptome (Romeuf, I., 2010) data, with the following questions:

- How can the transcript which corresponds to our identified protein be obtained?
- What is the ratio of significant correlations among expression profiles of proteome and transcriptome data?

We applied this linkage to enzymes of carbohydrate metabolism, to determine if it would provide complementary information for this functional category.

Chapter 2 (Article 1**): Proteomic analysis of peripheral layers during wheat (*Triticum aestivum* L.) grain development**

PROTEOMICS Volume 11, Issue 3, pages 371–379, No. 3 February 2011

Proteomics 2011, *11*, 371–379 DOI 10.1002/pmic.201000333 371

RESEARCH ARTICLE

Proteomic analysis of peripheral layers during wheat (*Triticum aestivum* L.) grain development

Ayesha Tasleem-Tahir[1], Isabelle Nadaud[1], Christine Girousse[1], Pierre Martre[1], Didier Marion[2] and Gérard Branlard[1]

[1] INRA, UMR 1095 GDEC-UBP, Clermont-Ferrand, France
[2] INRA, BIA, Nantes, France

Grains of hexaploid wheat, *Triticum aestivum* (cv. Récital), were collected at 15 stages of development, from anthesis to physiological maturity, 0–700°C days (degree days after anthesis). Two hundred and seven proteins of grain peripheral layers (inner pericarp, hyaline, testa and aleurone layer) were identified by 2-DE, MALDI-TOF MS and data mining, then were classified in 16 different functional categories. Study of the protein expression over time allowed identification of five main profiles and four distinct phases of development. Composite expression curves indicated that there was a shift from metabolic processes, translation, transcription and ATP interconversion towards storage and defence processes. Protein synthesis, protein turnover, signal transduction, membrane transport and biosynthesis of secondary metabolites were the mediating functions of this shift. A picture of the dynamic processes taking place in peripheral layers during grain development was obtained in this study. It should further help in the construction of proteome reference maps for the developing peripheral layers.

Received: May 26, 2010
Revised: October 13, 2010
Accepted: November 2, 2010

Keywords:
Aleurone layer / Expression profile / Grain development / Peripheral layers / Plant proteomics / Wheat

1 Introduction

Hexaploid wheat is one of the most important cereals for human nutrition. Understanding of the mechanisms taking place in wheat grain development will help improvement of qualitatively and quantitatively important traits. The endosperm in wheat [1] and maize [2] at different developmental stages has been studied by proteomics. Wheat bran layers also called peripheral layers (PLs) are composed of several cell layers: aleurone layer (AL), nucellus tegument or hyaline, testa (derived from inner ovary wall) and the pericarp layer. In PLs, several molecules of nutritional interest are concentrated especially vitamins, minerals [3, 4] and dietary fibers [5]; they are also rich in anti-oxidants [6]. The health benefits potentially related to this composition [7] have led to recommendation of whole grain consumption as a possible protection against the development of diet-related disorders such as cardiovascular disease, type 2 diabetes and against certain cancers [8]. However, PLs can also concentrate unwanted toxic pollutants such as mycotoxins, pesticide residues and heavy metals [9, 10]. It is therefore important to explore the formation of these tissues in relation with inter-specific and intra-specific variations of their composition and properties.

Since grain development is associated with massive changes in gene expression, any comparison between genotypes or environments is possible only by placing the results in a developmental context. Wheat grain outer layers have been studied by transcriptomics [11] and cytological tools [12] during grain development. As the protein content of the cells varies widely despite of the genome homology

Correspondence: Dr. Gérard Branlard, INRA UMR 1095 GDEC-UBP, 234 avenue du Brézet, 63100 Clermont-Ferrand, France
E-mail: branlard@clermont.inra.fr
Fax: +33-473624453

Abbreviations: AL, aleurone layer; **BSM,** Biosynthesis of Secondary Metabolites; **Cd,** degree days after anthesis; **CM,** carbohydrate metabolism; **EM,** energy metabolism; **HCA,** hierarchical clustering analysis; **PS,** protein synthesis; **PL,** peripheral layers; **RNP,** ribonucleoprotein; **SP,** storage protein; **UP,** unknown or hypothetical

[13], proteome data of PLs is important to provide complementary information. Separate tissue analysis of pericarp, intermediate and AL has been reported but only at grain maturity [14, 15].

The objective of the present study was to explore the proteome of wheat grain PLs, i.e. to identify proteins present in these complex tissues of diploid and triploid origin, and to analyse their quantitative and qualitative variations during grain development. The biochemical functions of the proteins identified made it possible to define the changes in major metabolic pathways that take place in PLs during grain development.

2 Materials and methods

2.1 Plant material

Analyses were performed on PLs isolated from grains of hexaploid winter wheat (*Triticum aestivum*) cv Récital. Seeds were sown in containers at INRA, Clermont-Ferrand, France. We used a 2:1 mixture of black soil: peat moss and sand 15%, enriched with substantive fertilizers (superphosphate 25% and potassium chloride 60%). Plants were placed under natural conditions in spring and were transferred to greenhouse 2 wk before anthesis (in natural light conditions). They were irrigated each week with tap water. Grains from the middle of the ears were tagged at the date of anthesis. Air temperature close to the ears was recorded and varied between 13.9 and 25.1°C. Daily mean air temperature was calculated, and the sum of mean temperatures was used to follow the developmental stage in thermal time (°Cd, degree days after anthesis). Wheat grains were harvested every 50°Cd from anthesis (0°Cd) to physiological maturity (700°Cd) and were stored at −80°C until dissection.

2.2 Separation of PLs

Dissection was performed on 25–50 grains, depending on the stage of development. The stages 0 and 59°Cd were exempted from dissection (as grains were very small and tissues were not fully differentiated) and whole grain proteins were used for analysis. Dissection was carried out under laminar air flow using binocular for grains of 109–700°Cd. First embryo and brush parts were removed. Then the dorsal part of the grain was incised with a scalpel and the outer pericarp was removed using tweezers. On the ventral part of seed, the crease was discarded by cutting its two longitudinal sides. The grain was opened in water containing 0.2% anti-protease cocktail, and the AL was scraped gently to remove the starchy endosperm.

Finally, the PLs composed of AL, hyaline layer, testa and inner pericarp were pooled in Eppendorf tubes after removing the water. They were weighed and then placed in liquid nitrogen. For each developmental stage, PLs were ground separately using pestle and mortar. The resulting powder was stored at −80°C until analysis.

2.3 Protein extraction

Proteins were extracted with cold extraction solution containing glacial acetone, 0.07% β-mercaptoethanol and 0.34% anti-protease cocktail, for 48 h at −20°C. Samples were centrifuged at 15 000 rpm at 0°C for 30 min. This was repeated until all the green colour of the pellet disappeared. The pellets were dried after removing the supernatant and then dissolved in buffer (7 M urea, 4% CHAPS, 2 M thiourea, 1% DTT and 1% IPG 3–10 buffer (GE Healthcare, Uppsala, Sweden). After 90-min incubation at room temperature, the extract was sonicated twice (15 W for 5 s) and left another 30 min incubation at room temperature. The mixture was then centrifuged for 5 min at 10 000 rpm at room temperature. The supernatant was recovered and protein concentration was measured by Bradford method [16]. The protein extracts were stored at −80°C until electrophoresis.

2.4 Two-dimensional electrophoresis

For characterisation of each stage, two biological extracts with two replicates per extract were used. IEF was performed using the IPGPhor II apparatus (GE Healthcare) on 24 cm Immobiline dry strips of 3–10 non-linear pH gradients. Passive re-hydration was performed overnight in a solution containing 7 M urea, 2 M thiourea, 70 mM DTT, 1% IPG buffer (pH 3–10), 4% CHAPS, 0.34% anti-protease and 300 μg of the protein extract. IEF was carried out by applying a cumulative voltage of 90 kVh as follows: 0.15 kV (1 h), 2.6 kV (4 h), 7.5 kV (3 h), 18 kV (3 h), and 61.733 kV (7.43 h).

Following IEF, proteins were reduced for 15 min in an equilibration buffer containing 0.05 M Tris-HCl (pH 8.8), 6 M urea, 30% glycerol, 2% SDS and 1% DTT, followed by alkylation for 15 min in the same buffer containing 2.5% iodoacetamide instead of DTT. The second dimension was performed using SDS-PAGE gels (14% T, 2.1% C) sealed with 0.5% agarose in SDS buffer on Ettan DALT II apparatus (GE Healthcare). The migration conditions were 5 W/gel for first 30 min then 10 W/gel until the exit of the dye front. Gels were stained using CBB G250 (Sigma, St. Louis, MO, USA) according to Rabilloud [17].

2.5 Image and statistical analysis

Gel images of 300 dpi and 16-bit greyscale pixel depth were obtained with G-800 (GE Healthcare) scanner and were analysed using SameSpots v3.2 (Nonlinear Dynamics, Newcastle, UK). Proteins with *p*-value of ANOVA and with

q-value (measure of false positives in data) less than 0.05 were considered significant. Normalised volume values were used for statistical tests. Principal component analysis (PCA) was performed on the set of gel images to detect possible outliers in the data. Hierarchical clustering analysis (HCA) was computed by using PermutMatrix v1.9.3 (LIRMM, Montpellier, France, http://www.lirmm.fr/~caraux/PermutMatrix/) according to Euclidean distance and Ward's method.

2.6 Protein identification

The proteins were identified using MALDI-TOF MS. Protein spots (313) were excised from gels (300 μg of protein extract/gel) and were destained with a solution containing 25 mM NH_4HCO_3, 5% ACN for 30 min and 25 mM NH_4HCO_3 50% ACN twice for 30 min. After dehydration in 100% ACN for 10 min, the spots were dried. Briefly, 150 ng of trypsin (V511, Promega, Madison, WI, USA), solution in 25 mM NH_4HCO_3 was added to the spots and digestion was performed at 37°C for 4–5 h. After centrifugation, peptides were extracted by adding 8 μL of ACN. For MALDI-TOF MS, 1 μL of peptides was loaded directly onto the MALDI target. The matrix solution (5 mg/mL CHCA in 50% ACN/0.1% TFA) was added immediately. After drying at room temperature, a Voyager DE-Pro model of MALDI-TOF mass spectrometer (Perseptive BioSystems, Framingham, MA, USA) was used in positive-ion reflector mode for PMF. External calibration was performed with a standard peptide solution (Proteomix C002, LaserBio Labs, Sophia-Antipolis, France). Internal calibration was performed using peptides resulting from auto-digestion of porcine trypsin.

Monoisotopic peptide masses were compared with those from Poaceae (15/09/2008, 20 903 136 sequences), Viridiplantae (13/01/2009, 965 593 sequences) and Wheat-EST (05/09/2008, 6 255 084 sequences) database using MASCOT (http://www.matrixscience.com). The Wheat-EST databank was constructed using the SRS (Sequence Retrieval System) from EBI (http://srs.ebi.ac.uk/) using a home-made PERL program.

The following parameters were considered for the searches: a maximum fragment ion mass tolerance of ±25 ppm, a maximum of one missed cleavage, partial methionine oxidation and partial carbamidomethylation of cysteine. If the MASCOT score was statistically significant ($p<0.05$), the protein was considered valid. EST similarity with other proteins can be visualised through blast at NCBI (National Center for Biotechnology Information, USA). These were then classified according to KEGG PATHWAY database (http://www.genome.jp/kegg/pathway.html) and gene ontology. We established composite expression profiles for each functional category by summing normalised volumes of all the protein spots (*s*) representative of that category for each developmental stage (*n*) and then the mean value for four replicates was expressed in μg/PLs of grain (Fig. 4).

$$Y = \frac{1}{4} \times \sum_{1}^{S} \%\text{volume} \times \frac{\text{quantity of protein extracted}_n}{100\,\text{mg of fresh mass of PLs}} \times \text{mean weight of PLs}_n$$

3 Results

3.1 Gel image analysis

Fifteen developmental stages were analysed using 2-DE. In our analysis, hand dissection adopted for separation of PLs [14] was found to be efficient, since no gluten proteins, characteristic of the starchy endosperm, were observed [18].

Gels were compared with a reference gel-denoted supermaster. This gel was obtained by combining the three different protein extracts: early, mid and later stages of development (153, 455 and 700°Cd, respectively). Spots were in the range of 13.3–82.4 kDa and p*I* of 4.3–9.1. At early stages of development, the majority of protein spots were recovered in the acidic zone of the gel, whereas at later stages more proteins were present in the basic zone of the gel (Fig. 1, Supporting Information Fig. 1).

The number of spots that appeared and disappeared between two adjacent stages and the number of common spots are shown in Fig. 2. These curves revealed four major periods of grain development: (i) cell division, (ii) cell differentiation, (iii) grain filling and (iv) grain dehydration corresponding to 0–59, 109–247, 295–455 and 501–700°Cd, respectively. The overall number of spots decreased towards maturity. Eight hundred and six significant protein spots based on the values of p and $q<0.05$ and power analysis value >0.8 were selected for Progenesis Stat analysis.

In PCA results, PC1 (first principal component) and PC2 (second principal component) explained 54.83 and 10.92% of the total variance (Supporting Information Fig. 2). The group of spots on the left corresponded to the proteins at early stages of development (0–247°Cd) and the group on the right side to the proteins present at later stages of development (295–700°Cd).

3.2 HCA

Two-way HCA was performed using 806 spots that were significant to group the proteins, showing similar expression profiles (Supporting Information Fig. 3). Five main expression profiles during development could be distinguished: (i) *Very early accumulation cluster*, with highest accumulation between 0 and 59°Cd, (ii) *Early accumulation cluster*, accumulation varies between early stages up to 455°Cd with peak at 295°Cd, (iii) *Early-mid accumulation cluster*, with accumu-

lation that shows a linear decrease from 204 to 501°Cd, (iv) *Mid-accumulation cluster*; with peak between 109 and 397°Cd, (v) *Late accumulation cluster*; with highest accumulation at later stages of development, mostly 350–700°Cd. The corresponding profiles encompassed 98, 137, 128, 54 and 268 protein spots, respectively. In these profiles, we found 21 spots specific to 153°Cd, 6 specific to 153, 204 and 247°Cd, and 46 specific to 658 and 700°Cd. There were 121 spots with irregular profile (indicated by X on Supporting Information Fig. 3). Four phases of grain development could be recognised by dendrogram of columns: 1(a) 0–59°Cd; 1(b) 109–247°Cd; 2(a) 295–455°Cd; 2(b) 501–700°Cd.

3.3 Protein classification

Among 313 protein spots subjected to MS analysis and database interrogation, 207 were identified (66.1% of protein

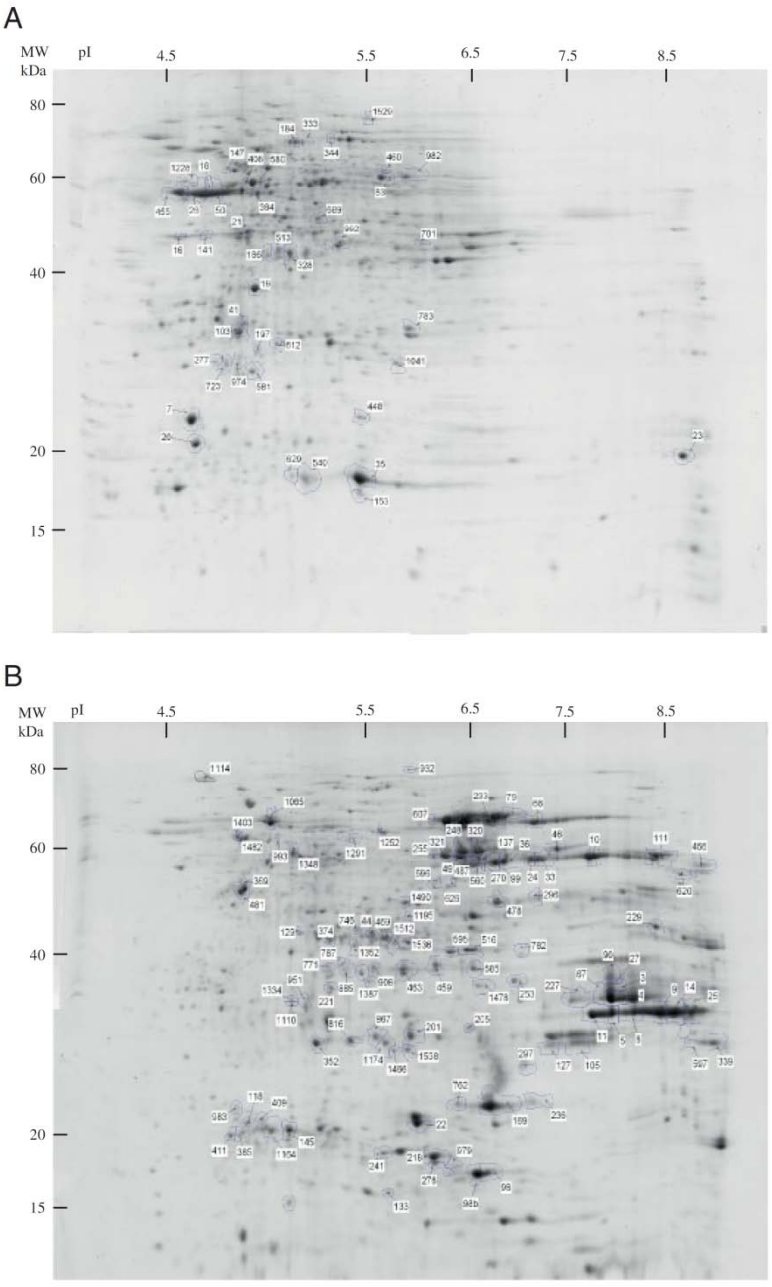


Figure 1. Proteins separated by 2-DE, numbered spots were identified (Supporting Information Table 1). (A) 153°Cd and (B) 700°Cd. Spot abundance shifts gradually from acidic to basic zone with kernel development.

Proteomics 2011, *11*, 371–379

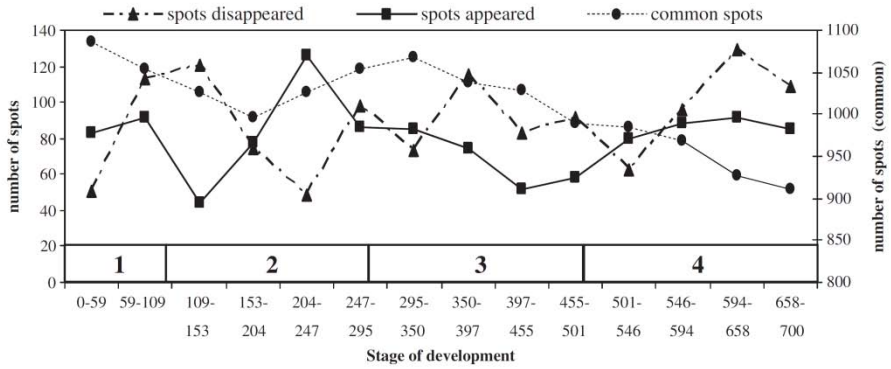


Figure 2. Number of spots appearing and disappearing between two adjacent stages and their common number of spots. We can observe four main phases of grain development according to spot number: (1) 0–59°Cd, (2) 109–247°Cd, (3) 295–455°Cd and (4) 501–700°Cd.

submitted) including proteins of profiles 2 to 5 (Supporting Information Table 1). After classification of proteins in their corresponding functional categories, the dominant functional class throughout development was metabolism with 32% of total identified proteins. The second largest functional class was of storage proteins (SP) 25% followed by stress defence 17%. Other proteins identified were genetic information and processing (GIP) 12%, unknown or hypothetical (UP) 6%, environmental information and processing (EIP) 3%, ATP inter-conversion proteins 3% and biosynthesis of secondary metabolites (BSM) 2%.

Among profiles (Fig. 3), the dominant functions of profile 2 were carbohydrate metabolism (CM) (glycolysis, starch and sucrose synthesis) with 18% proteins and protein synthesis (PS) with 16% (HSP 70, protein disulphide isomerase and elongation factor 1-γ). Profile 2 showed, however, the largest functional diversity with proteins involved in translation and UP (11% each), energy metabolism (EM) (carbon fixation, 8%), metabolism of other amino acids, transcription, and ATP inter-conversion proteins (6% each), protein turnover, protein folding, signal transduction, BSM, miscellaneous stress defence and amino acid metabolism (3% each).

In profile 3, the proteins involved in CM 25% and PS 16% (HSP 70) were the prominent classes. In profile 4, highest functional class was EM 48% (with chlorophyll-binding proteins and proteins involved in Calvin cycle-like RubisCO and transketolase) (Supporting Information Table 2) followed by CM 16% (with enzymes involved in glycolysis). Profile 5 included 45% of SP (mainly globulins) and 26% defence proteins (mainly xylanase inhibitor and chitinases) as major functional proteins.

3.4 Composite expression profiles of functional categories

Composite expression analysis was performed to examine the expression tendency for proteins of a given functional category [2] at each of the 15 stages of development (Fig. 4). Relative abundance of proteins involved in metabolism (energy, carbohydrate, amino acid and lipid) displayed a peak of expression between 204 and 247°Cd, indicating that metabolic activity was curtailed as grain matured (Fig. 4A). EM proteins accumulated between 204 and 247°Cd as for proteins related to the chloroplast machinery (Fig. 4A4), whereas proteins involved in amino acid metabolism, lipid metabolism such as CM (Fig. 4A1–A3) did not show any significant trend as compared with EM. Proteins of lipid metabolism peaked at 397°Cd. A slight decrease of expression was noticed after 295 and 204°Cd for proteins of the CM and amino acid metabolism functional groups. ATP inter-conversion was a prominent feature between 109 and 350°Cd (Fig. 4B). Protein folding, signal transduction and PS were at their lowest level at 204°Cd (Fig. 4C, E, and F). Protein turnover attained a peak at 153°Cd (Fig. 4E). Transcription and translation trends decreased as grain matured (Fig. 4D). Membrane transport along with BSM showed its peak towards maturity (Fig. 4H). Preponderance of SP was shown at the later stages of development and their content reached a peak at 658°Cd (Fig. 4I). Seed defence-related proteins increased in PLs during the later stages of development (Fig. 4I and J).

4 Discussion

PLs of wheat are of maternal origin except for AL which results from double fertilisation. They are composed of several cellular layers and comprehensive tissue analysis is required to unravel their composition as, in spite of their nutritional importance, they are generally eliminated in wheat flour preparation.

The grain grows very rapidly after fertilisation and the cell number in the tegument is determined early, approximately 100°Cd [19]. Cell differentiation gives rise to specialised tissues such as aleurone at 100–135°Cd [20]. In the endosperm, meanwhile, cell multiplication occurs and the number of cells at the end of the division phase was correlated to final grain mass [21]. Maternal layers undergo compression from internal expansion of the endosperm and 80% of final grain size is determined early, at 280°Cd [20]. Proteins of the profile 4 that accumulate up to 397°Cd showed that the major proteins were those linked with EM

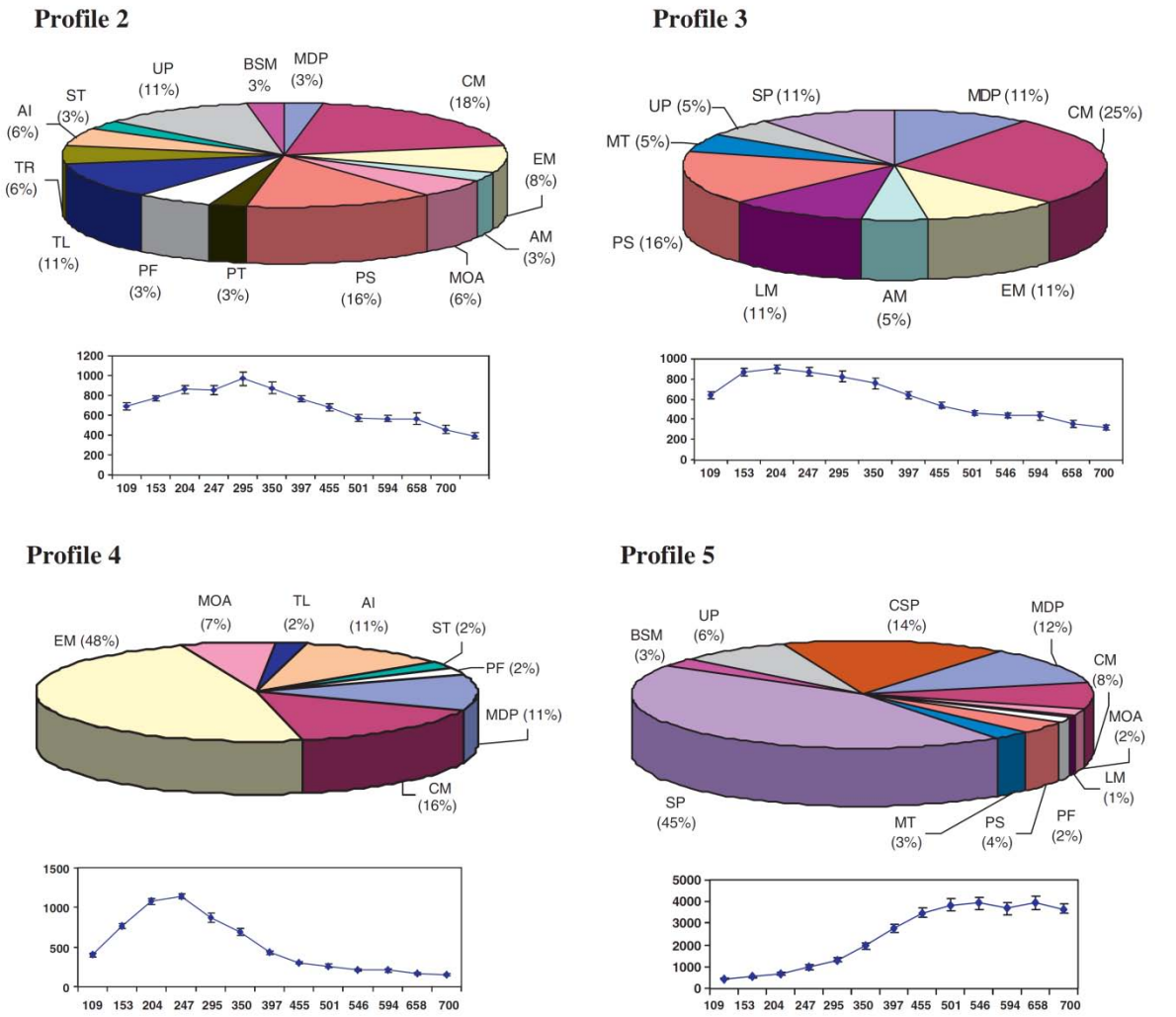


Figure 3. Functional classification of the identified proteins present in each of the four defined clusters. The percentage of identified spots included in each category is indicated in brackets. CM, carbohydrate metabolism; EM, energy metabolism; LM, lipid metabolism; AM, amino acid metabolism; MOA, metabolism of other amino acid; CSP, cell wall synthesis/related proteins; MDP, miscellaneous stress defence proteins; PS, protein synthesis; PT, protein turnover; PF, protein folding; TL, translation; TR, transcription; AI, ATP interconversion; ST, signal transduction; MT, membrane transport; BSM, biosynthesis of secondary metabolites; SP, storage proteins; UP, unknown or hypothetical proteins. Each curve below the pie chart showed the evolution of spots in corresponding cluster; stages of development are in *x*-axis and *y*-axis represents quantity of protein in μg/PLs of grain.

(in agreement with requirement for the functions described above). In accordance with photosynthetic activity of green pericarp, chlorophyll a-b binding protein and RubisCO enzyme were identified. It is noticeable that significant decrease of protein folding, PS and signal transduction occurred at 204°Cd (Fig. 4C, E, and F) that could be associated to the end of cell differentiation and thereafter increased probably with accumulation of SP (globulin) in AL (Fig. 4I). Metabolic processes were at their peak at 247°Cd.

Profile 2 showed maximum functional diversity with 13 different functions. This profile expressed a peak at 295°Cd and then decreased linearly with thermal time. CM was the largest category represented by this profile. Enzymes that functioned in citric acid cycle, glycolysis and sucrose synthesis were identified. Two ribonucleoproteins (RNP) were also identified in this profile. RNP complexes result from interaction of many different classes of RNA-binding proteins with various small non-coding RNA. They are actively involved in different aspects of cell metabolism,

Proteomics 2011, *11*, 371–379

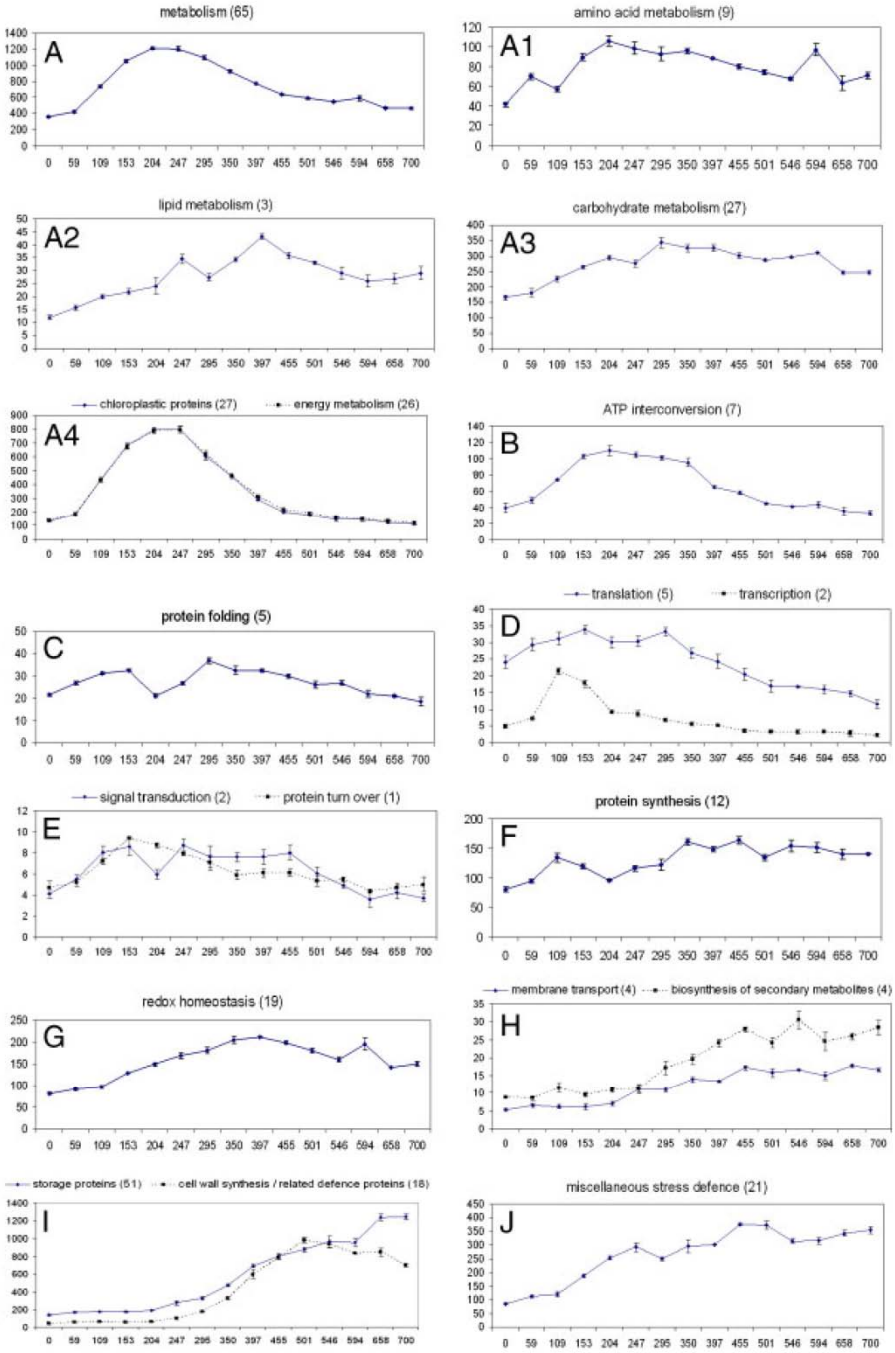


Figure 4. (A–J) Composite expression profiles where each point represents the mean of composite expression of proteins expressed in μg/PLs of grain (*y*-axis) at stages of development (*x*-axis) and the standard deviation ($n = 4$). The number of spots belonging to the same functional category is indicated in brackets.

such as DNA replication, expression of histone genes, regulation of transcription and translation [22]. In our analysis, sequence interrogation of RNP revealed that they were involved in post-transcriptional changes.

Late accumulation cluster (profile 5) was dominated by storage and defence-related proteins. The 51 SP identified in PLs together amounted to 1300 μg/PLs of grain at the end of grain development (700°Cd). Among these SP, globulins 2, 3, 3B and 3C were identified. These findings are consistent with the previous study on wheat endosperm, which showed that SP dominated late in development [1]. Three wheat globulin genes Glo-3A, Glo-3B and Glo-3C were identified as proteins potentially associated with the pathogenesis of T1D (Type 1 diabetes). Proteins expressed from Glo-3A and

Glo-3B had a molecular weight of 66.3 and 56.9 KDa, and had predicted p*I*s of 7.7 and 7.5, respectively. Immunofluorescence microscopy revealed the expression of Glo-3 specifically in the AL and the embryo of wheat grain [23]. Halford and Shewry [24] also found that 7S globulins from cereals are expressed predominantly in the embryo and AL. In our material, as we removed the embryo, the corresponding proteins were probably specific to AL. Since wheat globulins showed significant identity with other known or putative protein allergens, further research to determine their status as food allergens is still needed.

Oxygen level balance is a complex phenomenon in plants and mammals and is affected by endogenous and exogenous oxygen level. In the growing embryo, it is controlled by the balance between carbon fixation by photosynthesis and respiration, which depend on the developmental stage [25] and external light conditions of the plant [26]. As an oxygen map for wheat grain has not yet been developed, it is considered that it would be similar to that of barley given by Rolletschek [27]. Pericarp plays an important role in grain gas exchange during barley grain development. At 80–270°Cd, oxygen level is high, since the pericarp is still a living permeable tissue [28]. According to the barley oxygen map during storage and pre-storage stages, mean oxygen level declined in endosperm, resulting in decreased cell energy status. The same trend of EM was found for PLs in the present study. A transition of grain colour, from green to pale yellow, was observed between 397 and 455°Cd (Supporting Information Fig. 4) in parallel to a significant decrease in EM towards grain maturity (Fig. 4A4). In the late accumulation cluster, spots corresponding to the alcohol dehydrogenase and pyruvate decarboxylase were identified. The presence of these enzymes indicated an oxygen deprivation of the PLs, i.e. a response to an hypoxia status [29].

Photosynthesis is normally a major source of reactive oxygen species (ROS) in green tissue. Reduced oxygen is also responsible for the production of ROS that can damage cell components and can have a signalling role. To regulate their level in cells, numerous antioxidant defences are activated [30]. Consistently, in maize endosperm proteins involved in cell rescue, oxidative stress defence, pathogenesis and other abiotic defence proteins accumulate in the early stages of development. These proteins diminished with the fall in oxygen availability towards maturity [2]. Although enzymes involved in the detoxification of oxidative stress such as ascorbate peroxidase, glyoxalase and thioredoxin peroxidase were particularly present during the early stages of development mainly due to the photosynthetic activity in green pericarp, they can be identified in all the profiles including later accumulation stages. The hypoxia observed at the end of grain development is a source of oxidative stress [31] and the presence of enzymes involved in detoxifying ROS should help in maintaining the AL alive in contrast to cereal starchy endosperm that undergoes cell death towards end of grain development. However, at later stages (profile 5) besides, oxidative stress proteins, the cell wall-related proteins were dominant category in accordance with the previous reports [1, 2]. Xylanase inhibitor protein, a defence-related protein [32] with structural homology to family 18 chitinases [33] was prominent at these stages along with other chitinase and endo-chitinase proteins.

In summary, wheat grain PLs at early stages are the site of metabolic activity, photosynthesis and all the metabolic pathways linked to the activities of the chloroplast including ROS production and detoxification. PLs except AL became oriented towards cell death at 397–455°Cd. Cell death could be ROS mediated but the corresponding cell stress appears closely related to the synthesis of stress-related (defence) proteins. At later stages of development, PLs are dominated by the activity of aleurone cells as this is the only living tissue. They constitute a sophisticated defence structure for the developing embryo and endosperm together with abundant SP, and lipid proteins such as oleosine.

This is the first study exploring the proteome of PLs during whole wheat grain development. It should increase understanding of nutritional composition and its technological impacts. Linkage between proteomic and transcriptomic data will help to unravel changes in metabolism during grain development. In addition, it should increase understanding of responses to pathogen attack and enrich the knowledge of molecular mechanisms associated with cell wall lignification, particularly of testa and pericarp that have great influence on grain size determination. Detailed analyses could lead to the development of proteomic markers, which could be efficient for their tissue and stage specificity and would be helpful to overcome unwanted components of PLs in flour.

The authors acknowledge gratefully Christophe Chambon for MS analysis, Felicity Vear for critical English reading and Higher Education Commission, Pakistan for financial support.

The authors have declared no conflict of interest.

5 References

[1] Vensel, W. H., Tanaka, C. K., Cai, N., Wong, J. H. et al., Developmental changes in the metabolic protein profiles of wheat endosperm. *Proteomics* 2005, *5*, 1594–1611.

[2] Mechin, V., Thevenot, C., Le Guilloux, M., Prioul, J. L., Damerval, C., Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. *Plant Physiol.* 2007, *143*, 1203–1219.

[3] Antoine, C., Lullien-Pellerin, V., Abecassis, J., Rouau, X., Nutritional interest of the wheat seed aleurone layer. *Sci. Aliment.* 2002, *22*, 545–556.

[4] Amrein, T. M., Granicher, P., Arrigoni, E., Amado, R., In vitro digestibility and colonic fermentability of aleurone isolated from wheat bran. *Lebensm Wiss. Technol.* 2003, *36*, 451–460.

[5] Morrison, I. N., Kuo, J., Obrien, T. P., Histochemistry and fine-structure of developing wheat aleurone cells. *Planta* 1975, *123*, 105–116.

[6] Anson, N. M., Havenaar, R., Bast, A., Haenen, G., Antioxidant and anti-inflammatory capacity of bioaccessible compounds from wheat fractions after gastrointestinal digestion. *J. Cereal Sci.* 2010, *51*, 110–114.

[7] Liu, R. H., Whole grain phytochemicals and health. *J. Cereal Sci.* 2007, *46*, 207–219.

[8] Jones, J. A., Grain-based foods and health. *Cereal Food World* 2006, *51*, 108–113.

[9] Aureli, G., D'Egidio, M. G. Efficacy of debranning on lowering of deoxynivalenol (DON) level in manufacturing processes of durum wheat. *Tec. Mol.* 2007, *58*, 729–733.

[10] Fleurat-Lessard, F., Chaurand, M., Marchegay, G., Abecassis, J., Effects of processing on the distribution of pirimiphos-methyl residues in milling fractions of durum wheat. *J. Stored Prod. Res.*2007, *43*, 384–395.

[11] Wan, Y. F., Poole, R. L., Huttly, A. K., Toscano-Underwood, C. et al., Transcriptome analysis of grain development in hexaploid wheat. *Biomed. Chromatogr. Genomics* 2008, *9*, DOI: 10.1186/1471-2164-9-121.

[12] Beaugrand, J., Cronier, D., Thiebeau, P., Schreiber, L. et al., Structure, chemical composition, and xylanase degradation of external layers isolated from developing wheat grain. *J. Agric. Food Chem.* 2004, *52*, 7108–7117.

[13] Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L. et al., Proteomic survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 11969–11974.

[14] Laubin, B., Lullien-Pellerin, V., Nadaud, I., Gaillard-Martinie, B. et al., Isolation of the wheat aleurone layer for 2D electrophoresis and proteomics analysis. *J. Cereal Sci.* 2008, *48*, 709–714.

[15] Jerkovic, A., Kriegel, A. M., Bradner, J. R., Atwell, B. J. et al., Strategic distribution of protective proteins within bran layers of wheat protects the nutrient-rich endosperm. *Plant Physiol.* 2010, *152*, 1459–1470.

[16] Bradford, M. M., Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal. Biochem.* 1976, *72*, 248–254.

[17] Rabilloud, T., (Ed.), *Proteome research: Two-dimensional Gel Electrophoresis and Identification Methods*, Springer, Germany 2000, pp. 107–126.

[18] Gobaa, S., Bancel, E., Kleijer, G., Stamp, P., Branlard, G., Effect of the 1BL1RS translocation on the wheat endosperm, as revealed by proteomic analysis. *Proteomics* 2007, *7*, 4349–4357.

[19] Schnyder, H., Baum, U., Growth of the grain of wheat (*Triticum aestivum* L.). The relationship between water content and dry matter accumulation. *Eur. J. Agron.* 1992, *1*, 51–57.

[20] Nadaud, I., Girousse, C., Debiton, C. et al., Proteomic and morphological analysis of early stages of wheat grain development. *Proteomics* 2010, *10*, 2901–2910.

[21] Brocklehurst, P. A., Factors controlling grain weight in wheat. *Nature* 1977, *266*, 348–349.

[22] Glisovic, T., Bachorik, J. L., Yong, J., Dreyfuss, G., RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 2008, *582*, 1977–1986.

[23] Loit, E., Melnyk, C. W., MacFarlane, A. J., Scott, F. W., Altosaar, I., Identification of three wheat globulin genes by screening a *Triticum aestivum* BAC genomic library with cDNA from a diabetes-associated globulin. *Biomed. Chromatogr. Plant Biol.* 2009, *9*, DOI: 10.1186/1471-2229-9-93.

[24] Halford, N. G., Shewry, P. R., The structure and expression of cereal storage protein genes. *Plant Cell Monogr.* 2007, *8*, 195–218.

[25] Goffman, F. D., Alonso, A. P., Schwender, J., Shachar-Hill, Y., Ohlrogge, J. B., Light enables a very high efficiency of carbon storage in developing embryos of rapeseed. *Plant Physiol.* 2005, *138*, 2269–2279.

[26] Borisjuk, L., Rolletschek, H., The oxygen status of the developing seed. *New Phytol.* 2009, *182*, 17–30.

[27] Rolletschek, H., Weschke, W., Weber, H., Wobus, U., Borisjuk, L., Energy state and its control on seed development: starch accumulation is associated with high ATP and steep oxygen gradients within barley grains. *J. Exp. Bot.* 2004, *55*, 1351–1359.

[28] Rolletschek, H., Koch, K., Wobus, U., Borisjuk, L., Positional cues for the starch/lipid balance in maize kernels and resource partitioning to the embryo. *Plant J.* 2005, *42*, 69–83.

[29] Fukao, T., Bailey-Serres, J., Plant responses to hypoxia – is survival a balancing act? *Trends Plant Sci.* 2004, *9*, 449–456.

[30] Kovtun, Y., Chiu, W. L., Tena, G., Sheen, J., Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 2940–2945.

[31] Blokhina, O., Virolainen, E., Fagerstedt, K. V., Antioxidants, oxidative damage and oxygen deprivation stress: a review. *Ann. Bot (Lond.)* 2003, *91*, 179–194.

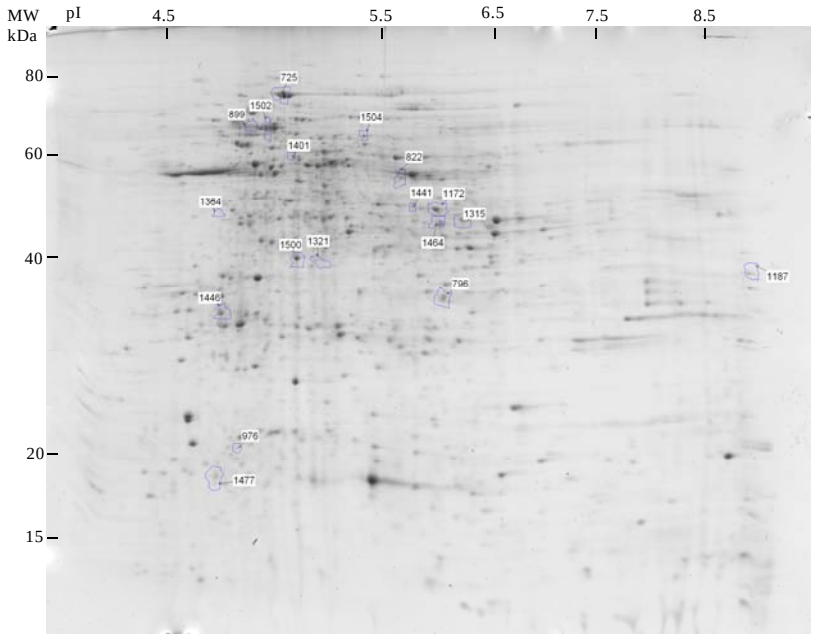
[32] Juge, N., Payan, F., Williamson, G., XIP-I, a xylanase inhibitor protein from wheat: a novel protein function. *Biochim Biophys. Acta Proteins Proteomics* 2004, *1696*, 203–211.

[33] Payan, F., Flatman, R., Porciero, S., Williamson, G. et al., Structural analysis of xylanase inhibitor protein I (XIP-I), a proteinaceous xylanase inhibitor from wheat (*Triticum aestivum*, var. Soisson). *Biochem. J.* 2003, *372*, 399–405.

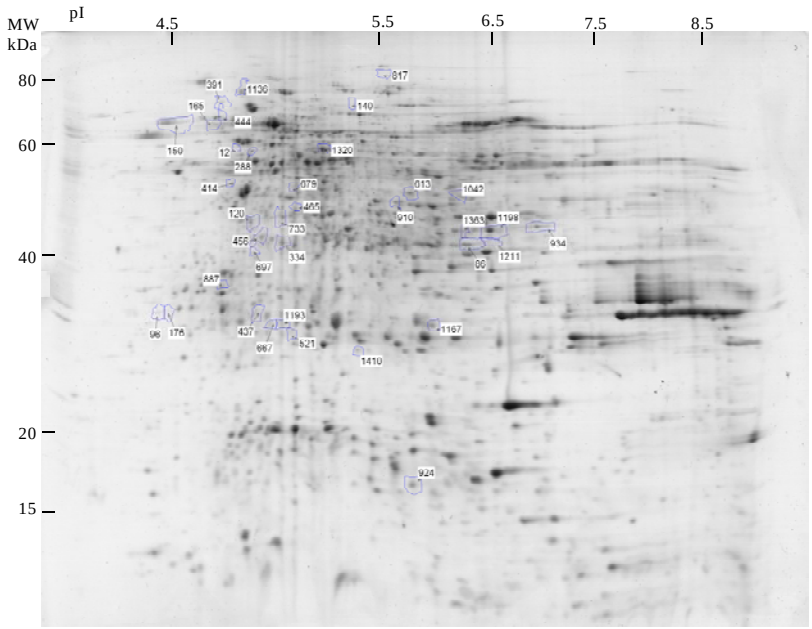
Additional Data: SuppInfo Table 1 and Table 5 are presented in Annex 1.

SupplInfo 1: Proteins separated by 2-DE, numbered spots were identified (table 1). a) 295°C ; b) 455°Cd. During these stages spots abundance shifts gradually from acidic to central zone and then centre to basic side of the gel.

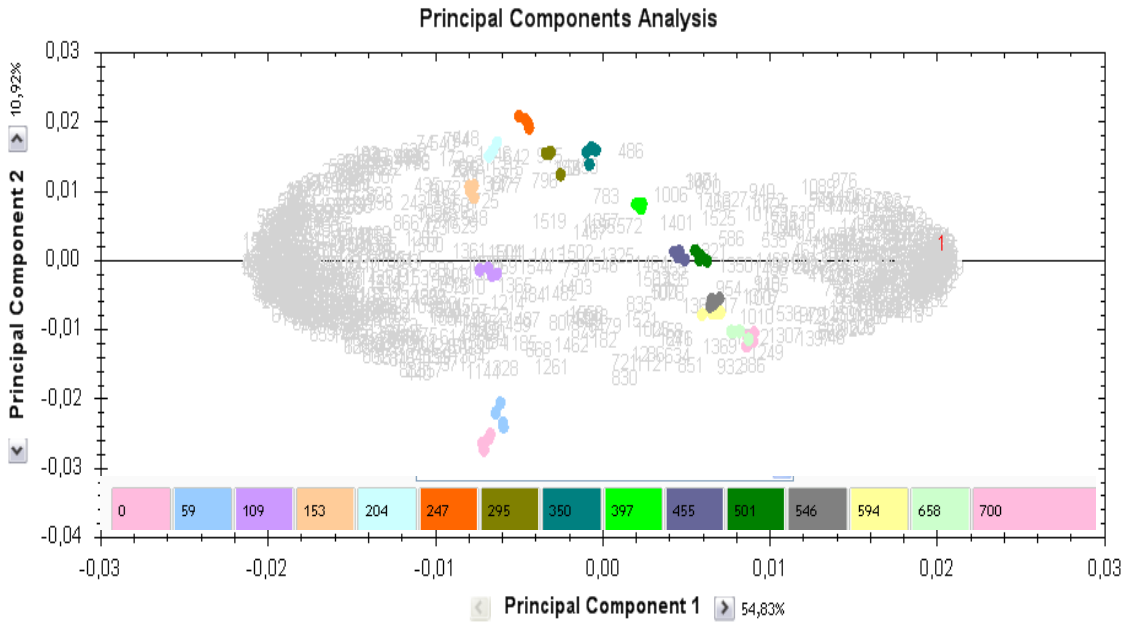
a



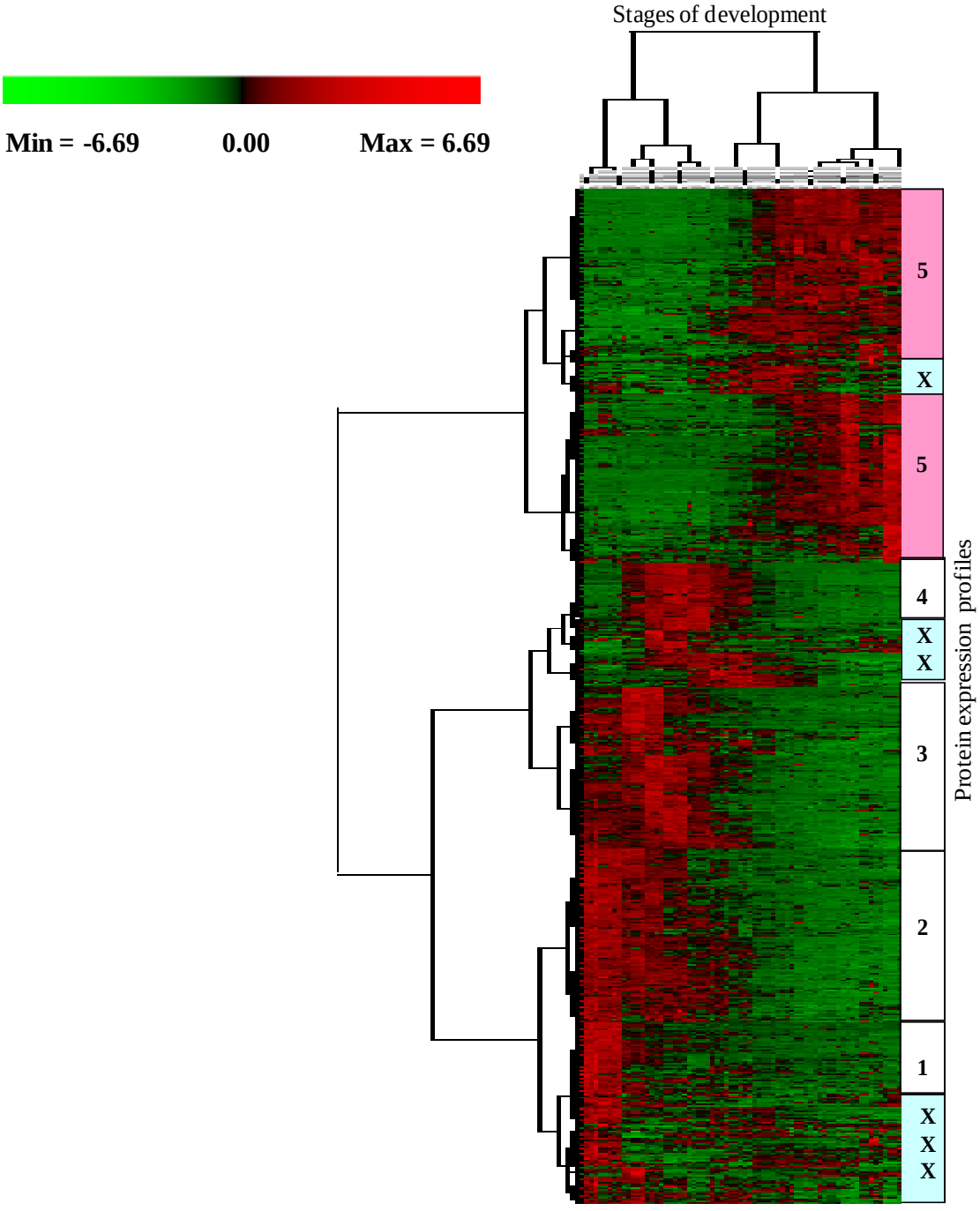
b



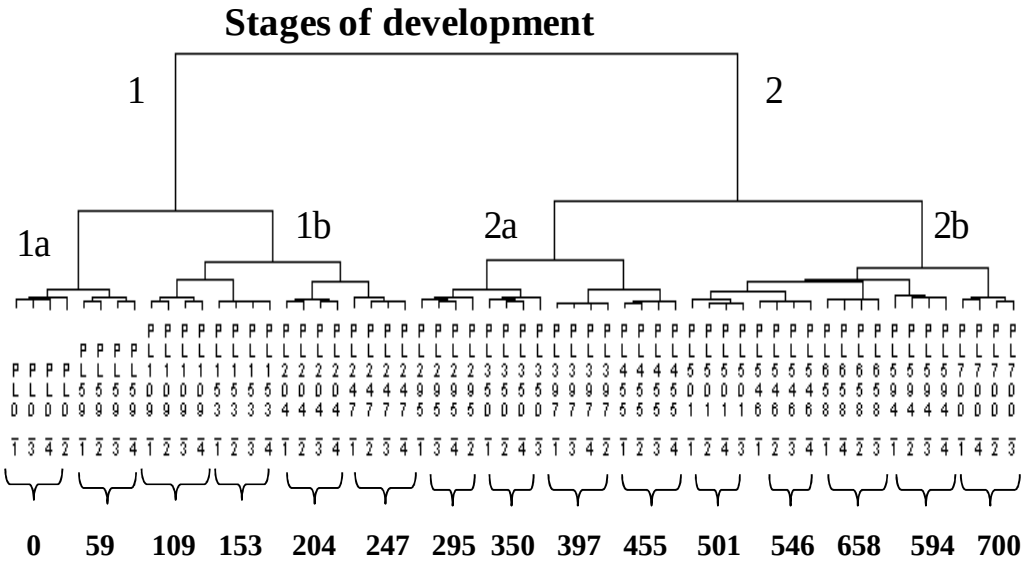
SupplInfo 2: Principal Component Analysis (15stages of development)



SupplInfo 3: Accumulation profiles by clustering of 806 significant spots during 15 stages of development. Protein expression profiles were obtained by clustering of protein spots, each horizontal line represents a protein spot. **X** (121 spots) represents common or irregular profiles, five main profiles were: **1** (98 spots) *Very early accumulation cluster;*
2 (137 spots) *Early accumulation cluster;* **3** (128 spots) *Early-mid accumulation cluster*;
4 (54 spots) *Mid-late accumulation cluster*; **5** (268 spots) *Late accumulation cluster*.



SupplInfo 3: Clustering revealed four main phases of development: 1a) 0-59°Cd; 1b) 109-247°Cd; 2a) 295-455°Cd; 2b) 501-700°Cd. Each vertical line (column) represents a gel.



SupplInfo 4: Grains of Récital at fifteen stages of development, color change between 397-455°C.

Chapter 3 (Article 2**): Expression profiling of starchy endosperm metabolic proteins at twenty-one stages of wheat grain development**

Journal of Proteome Research, 2012, 11 (5), pp 2754–2773

Journal of proteome research

Article

pubs.acs.org/jpr

Expression Profiling of Starchy Endosperm Metabolic Proteins at 21 Stages of Wheat Grain Development

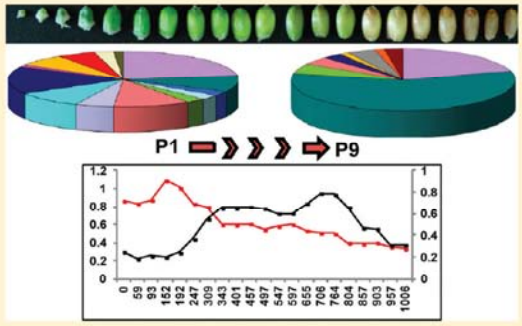
Ayesha Tasleem-Tahir,[†] Isabelle Nadaud,[†] Christophe Chambon,[‡] and Gérard Branlard*[,†]

[†]INRA, UMR 1095 GDEC-UBP, 234 avenue du Brézet, F-63100 Clermont-Ferrand, France

[‡]INRA, QPA, Proteomic Plateforme, F-63122 Saint-Genès Champanelle, France

Supporting Information

ABSTRACT: Proteomic analysis of albumins and globulins (alg) present in starchy endosperm of wheat (*Triticum aestivum* cv Récital), at 21 stages of grain development, led to the identification of 487 proteins. Four main developmental phases of these metabolic proteins, with three subphases in phase three and two in phase four, were shown. Hierarchical cluster analysis revealed nine major expression profiles throughout grain development. Classification of identified proteins in 17 different biochemical functions provided a uniform picture of temporal coordination among cellular processes. Proteins involved in cell division, transcription/translation, ATP interconversion, protein synthesis, protein transport, along with amino acid, lipid, carbohydrate and nucleotide metabolisms were highly expressed in early and early mid stages of development. Protein folding, cytoskeleton, and storage proteins peaked during the middle of grain development, while in later stages stress/defense, folic acid metabolism, and protein turn over were the abundant functional categories. Detailed analysis of stress/defense enzymes revealed three different evolutionary profiles. A global map with their predicted subcellular localizations and placement in grain developmental scale was constructed. The present study of complete grain development enriched our knowledge on proteome expression of alg, successively from endosperm cell division and differentiation to programmed cell death.



KEYWORDS: *proteomics, grain development, albumins and globulins, subcellular localization, stress/defense, Triticum aestivum*

INTRODUCTION

Grain quality improvement is the basic objective of wheat breeding, as wheat grains are the most important food source for humans. The major constituents of wheat grain are starch (70−80% dry weight) and proteins (10−15% dry weight). Of total wheat grain proteins, 80% are prolamins, monomeric gliadins and polymeric glutenins, while the nonprolamin part, albumins and globulins (alg), accounts for 15−20%.[1] Prolamins are the major storage proteins and are well characterized using whole grain or flour extracts.[2−5] In contrast to prolamins, alg play important roles in cellular metabolism, development and responses to environment as many of them are enzymes or enzyme inhibitors that influence technological quality.[6,7] Essential amino acids such as aspartate, threonine, lysine, and tryptophan are more abundant in alg as compared to storage proteins. They have complex genetic control, and their role in grain quality is less well understood compared with prolamins.[8]

Changes in metabolite levels can also be coregulated with seed development. Seed development is genetically programmed and can be regulated and influenced by sugar and nitrogen signals.[9−11] Analysis of results in a developmental context helps to provide detailed knowledge of mechanisms involved in the maturation of targeted sample (grain/tissue). In recent years, proteomic studies have been made on seeds and seed tissue development in cereals[12−14] and noncereals.[15,16] Attempts have been made to characterize alg proteins. For example, Singh et al. reported 19 alg proteins by Edman sequencing,[17] Wong et al. found several thioredoxin targets,[18] Vensel et al. studied 254 alg proteins at two stages of development, and Gao et al. published 230 alg proteins related to wheat quality.[19,20]

However, this knowledge is too limited to follow alg and related metabolic changes precisely during whole kernel development. To this aim, we studied developmental events of starchy endosperm (a tissue of grain reserve accumulation for the developing embryo) alg from fertilization to maturity at 21 stages of wheat grain development. To complement the results already published, we focused on a very short temporal interval of grain development, nearly at each 50 °Cd (degree days after pollination). We analyzed quantitative and qualitative (presence/absence of protein spots) variations, and we established accumulation profiles of these nonabundant endosperm proteins. We also described interconnections of different biochemical events between adjacent and nonadjacent stages of

Received: November 7, 2011
Published: March 7, 2012

dx.doi.org/10.1021/pr201110d | J. Proteome Res. 2012, 11, 2754−2773

Table 1. Proteins of Starchy Endosperm during Wheat Grain Development (*Triticum aestivum* cv Récital) Identified by LC–MS/MS and Classified by Their Corresponding Metabolic Pathway and Biochemical Functions[a]

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
Profile 1				
Stress/ defense				
1841*	gi257696322	gi3688398	Ascorbate peroxidase	cyto
1876*	gi28192421		Dehydroascorbate reductase	cyto
2659	gi63021412		Salt tolerant protein	cyto
853*	gi684948		Catalase	perox
2094	gi63021412		Salt tolerant protein	cyto
1945*	gi125663927		Manganese superoxide dismutase	cyto
1700*	gi115446749	Q84VC8	Gamma hydroxybutyrate dehydrogenase-like protein	nucl
2151	gi451193		Wali7	cyto
885a	gi18449343		Putative aldehyde dehydrogenase WIS1	M
1837*	gi145326672		Lactoylglutathione lyase, putative	cyto
Transcription / translation				
1789	gi257738538	gi357122287	Predicted 40S ribosomal protein SA	cyto
1510	gi115449323	P17070	Proliferating cell nuclear antigen	nucl
1456	gi242035869		Hypothetical protein SORBIDRAFT_01g036580	cyto
1838	gi242047528		Hypothetical protein SORBIDRAFT_02g003760	M
Amino acid Metabolism				
1138	gi283806355		Aminotransferase	M
1300	gi40317416		Glutamine synthetase isoform GSr1	cyto
960	gi68655466		Putative S-adenosylhomocystein hydrolase 2	cyto
1299	gi40317416		Glutamine synthetase isoform GSr1	cyto
1315	gi257676137	gi68655495	Methionine synthase 1 enzyme	cyto
1030	gi296514140	gi195622500	Serine hydroxymethyltransferase	cyto
606	gi257676137	gi68655495	Methionine synthase 1 enzyme	cyto
623	gi257676137	gi68655495	Methionine synthase 1 enzyme	cyto
627	gi257676183	gi68655500	Methionine synthase 2 enzyme	cyto
628	gi257676183	gi68655500	Methionine synthase 2 enzyme	cyto
689	gi257635198	gi195649503	Phenylalanine ammonia-lyase	cyto
1234	gi195957723		Aspartate aminotransferase	cyto
Protein turn over				
1686	gi242081935	C5YH32	Proteasome subunit alpha type	cyto
1889	gi52548242		20S proteasome beta 7 subunit	cyto
1844	gi195611928		Proteasome subunit beta type 1	cyto
1925	gi52548238		20S proteasome beta 4 subunit	cyto
1833	gi115447473	Q6H852	Proteasome subunit alpha type	cyto
1805	gi117670154	gi50080305	Putative proteasome subunit alpha type 3	cyto
Protein synthesis				
744	gi226533868		Heat shock protein 70	M
1778	gi108706531		40S ribosomal protein SA, putative, expressed	cyto
1816	gi108706531		40S ribosomal protein SA, putative, expressed	cyto
701	gi2827002		HSP70	cyto
739	gi226533868		Heat shock protein 70	M
713	gi188011548		Heat shock protein 70	cyto
700	gi115486793	gi21664287	Heat shock protein 70	cyto
Protein Folding				
1484	gi47118046		Protein disulfide isomerase	ER
1654	gi1729864		T-complex protein 1 subunit epsilon	cyto
2142	gi13925737		Cyclophilin A-3	cyto
986	gi296524608	gi56606827	Calreticulin-like protein	ER
Carbohydrate metabolism				
798*	gi158701881		NADP-dependent malic enzyme	cyto
994*	gi291047856	gi20530127	Mitochondrial aldehyde dehydrogenase	M
615	gi227483057	gi28190676	Putative transketolase	P
892	gi296523708	gi162457852	Pyruvate decarboxylase isozyme 3	cyto
604	gi227483057	gi28190676	Putative transketolase	P
1113*	gi218401053	gi108708038	Fumarate hydratase 1	M
1908	gi122171022		Soluble inorganic pyrophosphatase	Cysk
1305	gi28172909		Cytosolic 3-phosphoglycerate kinase	cyto
735	gi18076790		Phosphoglucomutase	cyto
1310	gi2204226		Alpha-galactosidase	nucl
1183	gi242064362		Hypothetical protein SORBIDRAFT_04g006440	M
885b*	gi218196143	B8AYE1	Dihydrolipoamide dehydrogenase	M
1033	gi6136111		UTP--glucose-1-phosphate uridylyltransferase	cyto
1855	gi259703023	gi11124572	Triosephosphat-isomerase	cyto

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
1483	gi226316439		Fructose-bisphosphate aldolase	cyto
1436*	gi115465579	Q6F361	Malate dehydrogenase	M
1419	gi223018643		Fructose-bisphosphate aldolase	P
1418	gi223018643		Fructose-bisphosphate aldolase	P
1012	gi6136111		UTP--glucose-1-phosphate uridylyltransferase	cyto
1455*	gi195628708		Malate dehydrogenase	M
877	gi18449343		Putative aldehyde dehydrogenase WIS1	M
1444*	gi115465579	Q6F361	Malate dehydrogenase	M
2635	gi18181983		Myo-inositol-1-phosphate synthase	cyto
1200	gi288814543		Alpha amylase	Extr
778	gi32400802		Phosphoglycerate mutase	cyto
1489*	gi32400786		NADPH-dependent mannose 6-phosphate reductase	perox
1427	gi32400859		Aldolase	P
1413	gi32400859		Aldolase	P
2126	gi259703023	gi11124572	Triosephosphat-isomerase	cyto
775	gi32400802		Phosphoglycerate mutase	cyto
1020*	gi242089615	C5Z0L0	Dihydrolipoyl dehydrogenase	P
Cell Wall				
1319	gi50659026		UDP-D-glucuronate decarboxylase	P
ATP Interconversion				
978	gi227303166	gi525291	ATP synthase beta subunit	M
1441*	gi20302473		Ferredoxin-NADP(H) oxidoreductase	P
Cell division				
390	gi257716429	gi357114006	Predicted cell division control protein 48 homolog E-like	cyto
Protein Transport				
1847*	gi126583387		Ferritin	P
Lipid Metabolism				
1532*	gi115478314		Os09g0277800	P
1479*	gi115475922		Os08g0327400	P
1723*	gi296510409	gi195626012	NADH-cytochrome b5 reductase-like protein	M
Nucleotide metabolism				
1323	gi4582787		Adenosine kinase	cyto
1321	gi4582787		Adenosine kinase	cyto
Unknown protein function				
1456a	gi115452789		Os03g0327600	cyto
2180	gi132270		Rubber elongation factor protein	cyto
Profile 2				
Stress/ defense				
1811*	gi15808779		Ascorbate peroxidase	cyto
1007*	gi238802280		Putative Td650 protein	cyto
1782*	gi15808779		Ascorbate peroxidase	cyto
1013*	gi238802280		Putative Td650 protein	cyto
1195	gi296511991	gi4666287	Cytosolic monodehydroascorbate reductase	cyto
1000	gi238802280		Putative Td650 protein	cyto
583	gi556673		Heat shock protein	P
1281	gi242038719		Hypothetical protein SORBIDRAFT_01g013540	P
Transcription/ translation				
2186	gi226497596	gi195604208	40S ribosomal protein S12	cyto
751	gi222629159		Hypothetical protein OsJ_15377	Extr
1098	gi257726667	gi162458395	Translational initiation factor eIF-4A	cyto
1373	gi242075424	gi18391211	TIF3H1; translation initiation factor	cyto
1110	gi257726665	gi162462542	Translational initiation factor	cyto
1249	gi257738538	gi357122287	Predicted 40S ribosomal protein SA	cyto
Amino acid Metabolism				
213	gi222618703		Hypothetical protein OsJ_02282	P
1275	gi506383		Glutamate 1-semialdehyde aminotransferase	P
612*	gi60686892		Delta 1-pyrroline-5-carboxylate synthetase	cyto
616	gi60686892		Delta 1-pyrroline-5-carboxylate synthetase	cyto
624	gi257676183	gi68655500	Methionine synthase 2 enzyme	cyto
2009	gi259439698	gi40363759	Putative glycine-rich protein	nucl
1473	gi295421187	gi585032	Cysteine synthase	cyto
847	gi109940331	gi28912436	Acetohydroxyacid synthase	cyto
1324bis	gi164471780		Aspartate aminotransferase	cyto
1042	gi115487944	Q2QVC1	Argininosuccinate synthase	P

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
Protein turn over				
1216	gi296524722	gi225216858	26S protease regulatory subunit S10B	cyto
1055	gi296525618	gi293331089	26S protease regulatory subunit 6B	cyto
1451	gi259662487	gi357146909	Proteasome subunit alpha type-1-like	cyto
1537	gi115435850	Q9SDD1	26S proteasome regulatory particle non-ATPase subunit11	nucl
Protein synthesis				
1215	gi115473357	Q8H3I3	Putative 40S ribosomal protein	cyto
601	gi254211611		70 kDa heat shock protein	P
654	gi476003		HSP70	ER
602	gi254211611		70 kDa heat shock protein	P
656	gi115477547	Q6ZD35	Putative glycyl-tRNA synthetase	M
1754	gi170753		Initiation factor (iso)4F p28 subunit	cyto
658	gi242033449	C5WQD6	Hypothetical protein SORBIDRAFT_01g012680	M
428	gi4103152		Histidyl-tRNA synthetase	cyto
Protein Folding				
1114	gi299469382		Putative PDI-like protein	ER
1092	gi299469382		Putative PDI-like protein	ER
857	gi115459800	Q7FAT6	T-complex protein 1 subunit alpha	cyto
772	gi296087296	gi15232923	Putative chaperonin	cyto
Carbohydrate metabolism				
1067	gi1707930		Glucose-1-phosphate adenylyltransferase large subunit	P
1578	gi46805452		Putative inorganic pyrophosphatase	P
1077	gi148508784		Glyceraldehyde-3-phosphate dehydrogenase	M
370	gi222628767		Hypothetical protein OsJ_14584	M
1294	gi115447367	Q6K9N6	Succinyl-CoA synthetase beta chain	M
1370	gi125560179	A2YRB6	Transaldolase family protein	P
1106	gi224021585		Plastid ADP-glucose pyrophosphorylase small subunit	P
1072	gi226532024		Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	M
824	gi115467370	gi15221156	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
1044	gi1707930		Glucose-1-phosphate adenylyltransferase large subunit	P
810	gi115448277	Q6ZFT9	Putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain	cyto
1116	gi218401053	gi108708038	Fumarate hydratase 1	M
1324	gi298541521	gi129916	Phosphoglycerate kinase, cytosolic	cyto
779	gi162957175		NADP-dependent malic enzyme 1	P
823	gi115467370	gi15221156	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
848	gi115467370	gi15221156	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
2664*	gi229358240		Cytosolic malate dehydrogenase	cyto
803	gi259708196	F2CX32	Pyruvate kinase	cyto
Signal transduction				
1694	gi227473229	gi112684	14-3-3-like protein	cyto
1410	gi126517972		Serine/threonine-protein phosphatase PP2A-1 catalytic subunit	cyto
Cell wall				
1364	gi4158230		Amylogenin	cyto
ATP Interconversion				
1061	gi296525622		Os06g0192600	Cysk
899	gi81176509		ATP1	M
325	gi222612529		Hypothetical protein OsJ_30891	cyto
283	gi222623539		Hypothetical protein OsJ_08115	ER
Cell division				
1608b	gi257666590	gi149392635	Nucleosome chromatin assembly protein	cyto
415	gi257716429	gi357114006	Predicted cell division control protein 48 homolog E-like	cyto
Protein Transport				
1766	gi16903082		Small Ras-related GTP-binding protein	cyto
1608a	gi297829788	gi15230476	Nascent polypeptide-associated complex subunit alpha-like protein 1	cyto
1740	gi195623482		Coatomer subunit epsilon	cyto
290	gi34393573		Putative karyopherin-beta 3 variant	cyto
Lipid Metabolism				
1494	gi242036027		Hypothetical protein SORBIDRAFT_01g038160	M
Cytoskeleton				
1196	gi58533119		Actin	Cysk
1207	gi226858185		Actin	Cysk
Profile 3				
Stress/ defense				
2544	gi2130114		Trypsin inhibitor CMx precursor	ER

dx.doi.org/10.1021/pr201110d | J. Proteome Res. 2012, 11, 2754–2773

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
1907*	gi259017810		Dehydroascorbate reductase	cyto
1685*	gi257696322	gi3688398	Ascorbate peroxidase	cyto
2293	gi134034577		Monomeric alpha-amylase inhibitor	Extr
Transcription/translation				
391	gi115462779	Q6F353	Putative minichromosome maintenance family protein	nucl
2340	gi114145394		Glycine-rich RNA-binding protein	nucl
Amino acid Metabolism				
2636	gi68655466		Putative S-adenosylhomocystein hydrolase 2	cyto
799	gi242091437	gi195627844	Ketol-acid reductoisomerase	P
Protein turn over				
1806	gi212721808	B4FDY6	Proteasome subunit alpha type	cyto
1964	gi242064246	C5XWW6	Proteasome subunit beta type	nucl
1673	gi242081935	C5YH32	Proteasome subunit alpha type	cyto
1915	gi49388033		Proteasome subunit beta type 3	cyto
Carbohydrate metabolism				
1543	gi7579064		Cytosolic glyceraldehyde-3-phosphate dehydrogenase	cyto
1235	gi298549023	gi226499486	Isocitrate dehydrogenase2	cyto
1124	gi226506764	B4FRC9	Transaldolase 2	P
408	gi75225211		Putative aconitate hydratase	cyto
959	gi195619804		Enolase	cyto
936	gi195619804		Enolase	cyto
1352*	gi229358240		Cytosolic malate dehydrogenase	cyto
1286	gi115447367	Q6K9N6	Succinyl-CoA synthetase beta chain	M
1374*	gi229358240		Cytosolic malate dehydrogenase	cyto
2549	gi210063883		Putative glucose-1-phosphate adenylyltransferase large subunit 1	P
943	gi115455455	gi195623986	UDP-glucose 6-dehydrogenase	ER
ATP Interconversion				
2283b	gi56784991		Putative ATP synthase beta subunit	M
1262	gi108925853		Vacuolar proton-ATPase C subunit	cyto
984	gi227303166	gi525291	ATP synthase beta subunit	M
2285	gi9652119		Nucleoside diphosphate kinase	cyto
2001	gi242043846	gi226507194	ATP synthase D chain, mitochondrial	M
Cell division				
1011	gi1556446		Alpha tubulin	cyto
1010	gi4098272		Alpha tubulin	cyto
1022	gi1556446		Alpha tubulin	cyto
Protein Transport				
1916	gi165973134	Q5XUV1	ADP-ribosylation factor	M
Lipid Metabolism				
1499	gi242036027		Hypothetical protein SORBIDRAFT_01g038160	M
1339	gi257342281	gi75247720	Stearoyl-ACP desaturase	P
Unknown protein function				
1378	gi226492599		Hypothetical protein LOC100277067	cyto
1845	gi132270		Rubber elongation factor protein	cyto
Profile 4				
Stress/ defense				
1794*	gi15808779		Ascorbate peroxidase	cyto
995*	gi684946		Catalase	perox
2264	gi39578552		Alpha amylase inhibitor CM3	Extr
1865*	gi15808779		Ascorbate peroxidase	cyto
1675*	gi15808779		Ascorbate peroxidase	cyto
1868	gi90959771		Multidomain cystatin	ER
413	gi296529766	gi110623251	Heat shock protein 101	cyto
1784bis	gi283480515		Tri a Bd 27K	Extr
Transcription/translation				
695a	gi115475838		Os08g0314800	cyto
695b	gi242065238	gi32492578	RNA binding protein	nucl
Amino acid Metabolism				
1035	gi1703227		Alanine aminotransferase 2	cyto
441	gi218401616	gi226502106	Glutaminyl-tRNA synthetase	cyto
880*	gi242091437	gi195627844	Ketol-acid reductoisomerase	P
611	gi257676137	gi68655495	Methionine synthase 1 enzyme	cyto

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
Protein turn over				
1208	gi296524722	gi225216858	26S protease regulatory subunit S10B	cyto
667	gi115450022	Q6K9T1	Oligopeptidase A-like	P
1059	gi115466876	gi6652878	26S proteasome AAA-ATPase subunit RPT1a	cyto
Protein synthesis				
680	gi476003		HSP70	ER
720	gi2827002		HSP70	cyto
653	gi476003		HSP70	ER
666	gi476003		HSP70	ER
Protein Folding				
1860	gi42493199		Cyclophilin A	cyto
980	gi77554944		Bifunctional aminoacyl-tRNA synthetase, putative	cyto
916	gi296524604	Q7Y140	Calreticulin	ER
930	gi296524604	Q7Y140	Calreticulin	ER
Carbohydrate metabolism				
401	gi75225211		Putative aconitate hydratase	cyto
890	gi91694277		Glucose-6-phosphate isomerase	cyto
966	gi1707923		Glucose-1-phosphate adenylyltransferase large subunit 1	P
2657*	gi229358240		Cytosolic malate dehydrogenase	cyto
1376*	gi229358240		Cytosolic malate dehydrogenase	cyto
828	gi115467370	gi15221156	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
998	gi20127139		Small subunit ADP glucose pyrophosphorylase	P
956	gi1707923		Glucose-1-phosphate adenylyltransferase large subunit 1	P
1784	gi257355184	O48556	Soluble inorganic pyrophosphatase	cyto
411	gi92429669		Putative aconitate hydratase 1	P
Cell wall				
1303	gi50659026		UDP-D-glucuronate decarboxylase	P
Lipid Metabolism				
1480	gi242036027		Hypothetical protein SORBIDRAFT_01g038160	M
Profile 5				
Stress/ defense				
2291	gi253783731		Alpha amylase inhibitor CM1	Extr
1255c*	gi119388709		Alcohol dehydrogenase ADH1	cyto
584	gi32765549		Cytosolic heat shock protein 90	cyto
1201	gi75313847		Serpin-Z2A	P
412	gi544242		Glucose-regulated protein 94 homolog	ER
1008*	gi238802280		Putative Td650 protein	cyto
901*	gi291047792	gi300087069	Aldehyde dehydrogenase 7b	perox
Transcription/translation				
1714	gi232033		Elongation factor 1-beta	cyto
1718	gi232033		Elongation factor 1-beta	cyto
2644	gi32400796		Acidic ribosomal protein	M
1095a	gi170776		Translation elongation factor 1 alpha-subunit	cyto
1095b	gi170776		Translation elongation factor 1 alpha-subunit	cyto
1103	gi257726667	gi162458395	Translational initiation factor eIF-4A	cyto
1956	gi257714428	gi195637912	Transcription factor BTF3	nucl
337	gi242065238	gi32492578	RNA binding protein	nucl
1934*	gi75246527		Translationally-controlled tumor protein homolog	cyto
409	gi115446385	Q6H4L2	Elongation factor 2	cyto
788	gi1737492		Poly(A)-binding protein	cyto
1081	gi170776		Translation elongation factor 1 alpha-subunit	cyto
1255a	gi170776		Translation elongation factor 1 alpha-subunit	cyto
Amino acid Metabolism				
212	gi222618703		Hypothetical protein OsJ_02282	P
1028	gi1703227		Alanine aminotransferase 2	cyto
1163	gi194268461		Chorismate synthase	P
1168	gi194268461		Chorismate synthase	P
888	gi242091437	gi195627844	Ketol-acid reductoisomerase	P
922	gi296514168	gi108862549	Serine hydroxymethyltransferase	cyto
442	gi218401616	gi226502106	Glutaminyl-tRNA synthetase	cyto
932	gi296514168	gi108862549	Serine hydroxymethyltransferase	cyto
1043	gi1703227		Alanine aminotransferase 2	cyto
2031	gi259439698	gi40363759	Putative glycine-rich protein	nucl

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
Protein synthesis				
585	gi4204859		Heat shock protein 80	cyto
710	gi476003		HSP70	ER
Protein Folding				
907	gi6671939		Putative T-complex protein 1, ETA subunit	cyto
834	gi47118046		Protein disulfide isomerase	ER
1301	gi299469378		Putative PDI-like protein	ER
2634	gi242032147		Hypothetical protein SORBIDRAFT_01g000380	P
2145a	gi13925737		Cyclophilin A-3	cyto
2145b	gi42493199		Cyclophilin A	cyto
Carbohydrate metabolism				
967	gi1707930		Glucose-1-phosphate adenylyltransferase large subunit	P
1798	gi259662377	gi226529672	Triosephosphate isomerase, cytosolic	cyto
1255b	gi298549023	gi226499486	Isocitrate dehydrogenase2	cyto
769	gi32400802		Phosphoglycerate mutase	cyto
568	gi401138		Sucrose synthase 1	M
540	gi11037530		Starch branching enyzyme 1	P
2650	gi148508784		Glyceraldehyde-3-phosphate dehydrogenase	M
569	gi401138		Sucrose synthase 1	M
1202*	gi242095836	gi195640660	Formate dehydrogenase 1	M
348	gi183211902		Plastid alpha-1,4-glucan phosphorylase	P
1191*	gi242095836	gi195640660	Formate dehydrogenase 1	M
546	gi11037530		Starch branching enzyme 1	P
344	gi183211902		Plastid alpha-1,4-glucan phosphorylase	P
326	gi222628767		Hypothetical protein OsJ_14584	M
1256*	gi242095836	gi195640660	Formate dehydrogenase 1	M
1259*	gi242095836	gi195640660	Formate dehydrogenase 1	M
774	gi115476012	Q84QT9	Putative pyrophosphate-dependent phosphofructokinase alpha subunit	cyto
1004	gi20127139		Small subunit ADP glucose pyrophosphorylase	P
Signal transduction				
1681	gi2492487		14-3-3-like protein B	cyto
1680	gi2492487		14-3-3-like protein B	cyto
1639	gi257664738	gi2266662	14-3-3 protein	cyto
1193	gi297851506		GTP binding protein	cyto
1643	gi257664738	gi2266662	14-3-3 protein	cyto
1644	gi257664738	gi2266662	14-3-3 protein	cyto
1634	gi257664738	gi2266662	14-3-3 protein	cyto
ATP Interconversion				
399	gi254256262	E9NQE5	Pyruvate orthophosphate dikinase 1	P
398	gi254256262	E9NQE5	Pyruvate orthophosphate dikinase 1	P
Cell division				
397	gi257716429	gi357114006	Predicted cell division control protein 48 homolog E-like	cyto
Protein Transport				
757	gi45357045		Coatomer alpha subunit	M
Folic acid metabolism				
717	gi242044850		Hypothetical protein SORBIDRAFT_02g026140	P
715	gi242044850		Hypothetical protein SORBIDRAFT_02g026140	P
730	gi242044850		Hypothetical protein SORBIDRAFT_02g026140	P
Profile 6				
Stress/ defense				
2205	gi115464233	gi195645676	USP family protein	cyto
Transcription/translation				
1954	gi41400293		S-like RNase	Extr
Amino acid Metabolism				
879*	gi291047652	gi21747870	Betaine-aldehyde dehydrogenase	cyto
Protein synthesis				
2162	gi296512518	gi186886530	16.8 kDa heat-shock protein	cyto
2161	gi296512688	gi123545	16.9 kDa class I heat shock protein 1	cyto
2165	gi296512518	gi186886530	16.8 kDa heat-shock protein	cyto
2166	gi296512797	gi195626536	17.4 kDa class I heat shock protein 3	cyto
2187	gi296512793	gi321266547	Heat shock protein 17	cyto
2135	gi296512688	gi123545	16.9 kDa class I heat shock protein 1	cyto
Protein Folding				
973	gi77554944		Bifunctional aminoacyl-tRNA synthetase, putative	cyto

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
Carbohydrate metabolism				
384	gi257665965	gi3341490	Phospoenolpyruvate carboxylase	cyto
876	gi298545815	gi4588609	Granule bound starch synthase precursor	P
400	gi75225211		Putative aconitate hydratase	cyto
1154	gi298541521	gi129916	Phosphoglycerate kinase, cytosolic	cyto
555	gi3393044		Sucrose synthase type 2	cyto
547	gi3393044		Sucrose synthase type 2	cyto
759	gi18076790		Phosphoglucomutase	cyto
552	gi401138		Sucrose synthase 1	M
1863	gi290875537		Putative carbonic anhydrase	P
Profile 7				
Stress/ defense				
1869*	gi34539782		1-Cys-peroxiredoxin	cyto
2339	gi54778521		0.19 dimeric alpha-amylase inhibitor	Extr
1887*	gi28192421		Dehydroascorbate reductase	cyto
1197	gi75313847		Serpin-Z2A	P
1189	gi224589266		Serpin 1	P
1274	gi224589268		Serpin 2	P
937	gi75313847		Serpin-Z2A	P
Transcription/translation				
761	gi222629159		Hypothetical protein OsJ_15377	Extr
760	gi222629159		Hypothetical protein OsJ_15377	Extr
Amino acid Metabolism				
1135	gi283806359		Alanine-glyoxylate aminotransferase	M
1345	gi164471780		Aspartate aminotransferase	cyto
2041	gi259439698	gi40363759	Putative glycine-rich protein	nucl
1435	gi212276289	B4F833	Diaminopimelate epimerase	P
Protein synthesis				
2224	gi75766428		Chain L,Negative Stain Em Reconstruction Of M.Tuberculosis Acr1(Hsp 16.3)	cyto
Protein Folding				
2637	gi242032147		Hypothetical protein SORBIDRAFT_01g000380	P
839	gi13925726		Protein disulfide isomerase 2 precursor	ER
Carbohydrate metabolism				
1472	gi32400764		Beta amylase	Extr
560	gi401138		Sucrose synthase 1	M
539	gi11037530		Starch branching enyzyme 1	P
554	gi3393044		Sucrose synthase type 2	cyto
570	gi3393044		Sucrose synthase type 2	cyto
571	gi3393044		Sucrose synthase type 2	cyto
576	gi3393044		Sucrose synthase type 2	cyto
573	gi3393044		Sucrose synthase type 2	cyto
561	gi3393044		Sucrose synthase type 2	cyto
918	gi91694277		Glucose-6-phosphate isomerase	cyto
718a	gi20259685		Beta-D-glucan exohydrolase	ER
1386	gi148508784		Glyceraldehyde-3-phosphate dehydrogenase	M
1318	gi253783729		Glyceraldehyde-3-phosphate dehydrogenase	cyto
1404*	gi125561648	A2YVI5	Malate dehydrogenase	P
1372	gi148508784		Glyceraldehyde-3-phosphate dehydrogenase	M
1458	gi34485587		Plastidic alpha 1,4-glucan phosphorylase 3	cyto
1332	gi226316439		Fructose-bisphosphate aldolase	cyto
1399*	gi229358240		Cytosolic malate dehydrogenase	cyto
920	gi91694277		Glucose-6-phosphate isomerase	cyto
2660*	gi229358240		Cytosolic malate dehydrogenase	cyto
2654*	gi229358240		Cytosolic malate dehydrogenase	cyto
Signal transduction				
1554	gi257672913	gi162459667	Annexin p33	cyto
Cell wall				
934	gi6175480		Xylose isomerase	ER
Protein Transport				
1971	gi115447377	Q6K1Q5	Glycolipid transfer protein-like	cyto
Storage Proteins				
1595	gi295853625		Avenin-like b	ER
1742	gi110341795		Globulin 1	ER

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
Folic acid metabolism				
1521	gi115589734		5,10-methylene-tetrahydrofolate dehydrogenase	cyto
Unknown protein function				
2553	gi132270		Rubber elongation factor protein	cyto
Profile 8				
Stress/ defense				
1406*	gi257710966	gi62765876	2-alkenal reductase	cyto
1614*	gi159895412		NADPH-dependent thioredoxin reductase isoform 2	cyto
2256	gi221855656		Alpha-amylase inhibitor CM16 subunit	Extr
1190	gi75313847		Serpin-Z2A	P
1351	gi224589270		Serpin 3	P
1330a	gi75279909		Serpin-Z2B	P
1679	gi51247633		Chain A, Crystal Structure Of Family 11 Xylanase In Complex With Inhibitor	Extr
1186	gi75313848		Serpin-Z1C	P
1221	gi75313848		Serpin-Z1C	P
1165	gi224589266		Serpin 1	P
2545	gi55669878		Chain B, Crystal Structure Of The Triticum Aestivum Xylanase Inhibitor-I	cyto
1265	gi224589268		Serpin 2	P
1175	gi75313848		Serpin-Z1C	P
1317	gi75313848		Serpin-Z1C	P
1757	gi51247633		Chain A, Crystal Structure Of Family 11 Xylanase In Complex With Inhibitor	Extr
1270	gi224589268		Serpin 2	P
1817	gi62465514		Class II chitinase	ER
2230	gi114215938		Dimeric alpha-amylase inhibitor	Extr
439	gi296529766	gi110623251	Heat shock protein 101	cyto
1615	gi226495167		Desiccation-related protein PCC13-62	ER
1645	gi145326672		Lactoylglutathione lyase, putative	cyto
1687	gi156186245		Xylanase inhibitor 725ACCN	ER
2214*	gi226897529		Superoxide dismutase	cyto
2270*	gi226897529		Superoxide dismutase	cyto
Amino acid Metabolism				
923	gi296514168	gi108862549	Serine hydroxymethyltransferase	cyto
Protein synthesis				
2266	gi75766428		Chain L,Negative Stain Em Reconstruction Of M.Tuberculosis Acr1(Hsp 16.3)	cyto
2206	gi75766428		Chain L,Negative Stain Em Reconstruction Of M.Tuberculosis Acr1(Hsp 16.3)	cyto
1944	gi157093720	gi18397757	Basic secretory protein family protein	ER
1227	gi2827002		HSP70	cyto
Protein Folding				
1589	gi299469378		Putative PDI-like protein	ER
870	gi2493650		60 kDa chaperonin subunit beta	cyto
862	gi2493650		60 kDa chaperonin subunit beta	cyto
Carbohydrate metabolism				
898	gi32400764		Beta amylase	Extr
909	gi38349539		Beta-amylase 1	Extr
1330b	gi298541521	gi129916	Phosphoglycerate kinase, cytosolic	cyto
1356	gi226316439		Fructose-bisphosphate aldolase	cyto
910	gi195619804		Enolase	cyto
1347	gi4158232		Reversibly glycosylated polypeptide	cyto
1366a	gi226316439		Fructose-bisphosphate aldolase	cyto
1366	gi226316439		Fructose-bisphosphate aldolase	cyto
1393*	gi229358240		Cytosolic malate dehydrogenase	cyto
1333	gi3646373		RGP1 protein	cyto
968	gi193073259		Beta-glucosidase	ER
1295	gi253783729		Glyceraldehyde-3-phosphate dehydrogenase	cyto
Signal transduction				
1667	gi227473229	gi112684	14-3-3-ike protein A	cyto
Cell wall				
1767	gi40363753		Putative caffeoyl CoA O-methyltransferase	cyto
1768	gi40363753		Putative caffeoyl CoA O-methyltransferase	cyto
Storage Proteins				
1691	gi133741924		Gamma gliadin	ER
Profile 9				
Stress/ defense				
2395*	gi226897529		Superoxide dismutase	cyto

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
1929*	gi257333180	Q8RW03	Glutathione transferase	cyto
1913	gi20257409		Thaumatin-like protein	ER
2350	gi54778521		0.19 dimeric alpha-amylase inhibitor	Extr
2480	gi134034647		Monomeric alpha-amylase inhibitor	Extr
2443	gi1588926		Pathogenesis-related protein	Extr
2350a	gi54778521		0.19 dimeric alpha-amylase inhibitor	Extr
2648	gi39578552		Alpha amylase inhibitor CM3	Extr
2649	gi39578552		Alpha amylase inhibitor CM3	Extr
2442	gi54778511		0.19 dimeric alpha-amylase inhibitor	Extr
2083	gi123975		Endogenous alpha-amylase/subtilisin inhibitor	cyto
2402	gi283465827		Putative alpha-amylase inhibitor CM2	Extr
2328	gi54778521		0.19 dimeric alpha-amylase inhibitor	Extr
2364	gi54778521		0.19 dimeric alpha-amylase inhibitor	Extr
2463	gi134034647		Monomeric alpha-amylase inhibitor	Extr
2423	gi134034647		Monomeric alpha-amylase inhibitor	Extr
1812	gi62465514		Class II chitinase	ER
1890	gi14164981		Thaumatin-like protein	ER
1931	gi20257409		Thaumatin-like protein	ER
1829	gi62465514		Class II chitinase	ER
1769	gi51247633		Chain A ,Crystal Structure Of Family 11 Xylanase In Complex With Inhibitor (XiP-I)	Extr
1504	gi113595		Aldose reductase	cyto
2421	gi54778511		0.19 dimeric alpha-amylase inhibitor	Extr
2475	gi134034647		Monomeric alpha-amylase inhibitor	Extr
2309	gi39578552		Alpha amylase inhibitor CM3	Extr
2411	gi54778511		0.19 dimeric alpha-amylase inhibitor	Extr
2387	gi114215876		Dimeric alpha-amylase inhibitor	Extr
2420	gi134034647		Monomeric alpha-amylase inhibitor	Extr
2358	gi114215938		Dimeric alpha-amylase inhibitor	Extr
2347	gi65993872		Dimeric alpha-amylase inhibitor	Extr
2333	gi114215938		Dimeric alpha-amylase inhibitor	Extr
1996	gi224589266		Serpin 1	P
1946*	gi125663927		Manganese superoxide dismutase	cyto
1943*	gi125663927		Manganese superoxide dismutase	cyto
1943a*	gi1621627		Manganese superoxide dismutase	M
1280	gi55669878		Chain B, Crystal Structure Of The Triticum Aestivum Xylanase Inhibitor-I	cyto
1185	gi75313848		Serpin-Z1C	P
1177	gi75279910		Serpin-Z1B	P
1156	gi75279909		Serpin-Z2B	P
1167	gi75279910		Serpin-Z1B	P
1565	gi495305		Chitinase	ER
2354	gi114215938		Dimeric alpha-amylase inhibitor	Extr
Amino acid Metabolism				
1278*	gi242073884	gi81686712	Glutamate dehydrogenase 2	M
1369	gi164471780		Aspartate aminotransferase	cyto
Protein turn over				
1900	gi66271071		Beta1 proteasome-1D	cyto
2404	gi257708815	P55857	Small ubiquitin-related modifier 1	cyto
2061	gi211906468		20S proteasome subunit alpha-1	cyto
1497	gi259662487	gi357146909	Proteasome subunit alpha type-1-like	cyto
Protein synthesis				
1260	gi2827002		HSP70	cyto
1279	gi476003		HSP70	ER
1213	gi476003		HSP70	ER
Protein Folding				
874	gi222446344		Protein disulfide isomerase	ER
Carbohydrate metabolism				
883	gi38349539		Beta-amylase 1	Extr
883a	gi38349539		Beta-amylase 1	Extr
882	gi38349539		Beta-amylase 1	Extr
837	gi32400764		Beta-amylase	Extr
836	gi32400764		Beta-amylase	Extr
2651	gi148508784		Glyceraldehyde-3-phosphate dehydrogenase	M

Table 1. continued

Spot ID	accession	accession after BlastP	Protein name	Predicted sub-cellular localization
1355	gi226316439		Fructose-bisphosphate aldolase	cyto
1365a	gi226316439		Fructose-bisphosphate aldolase	cyto
974	gi193073259		Beta-glucosidase	ER
965	gi193073259		Beta-glucosidase	ER
975	gi193073259		Beta-glucosidase	ER
1365b	gi226316439		Fructose-bisphosphate aldolase	cyto
1380	gi148508784		Glyceraldehyde-3-phosphate dehydrogenase	M
1415*	gi125561648	A2YVI5	Malate dehydrogenase	P
889	gi38349539		Beta-amylase 1	Extr
Storage Proteins				
2662	gi110341801		Globulin 1	ER
Folic acid metabolism				
782	gi242044850		Hypothetical protein SORBIDRAFT_02g026140	P
855	gi115589734		5,10-methylene-tetrahydrofolate dehydrogenase	cyto
Unknown protein function				
2646	gi132270		Rubber elongation factor protein	cyto
1950	gi132270		Rubber elongation factor protein	cyto
2383	gi132270		Rubber elongation factor protein	cyto

[a]Spot IDs with asterix (*) represent proteins of redox homeostasis. Spot ID; spot number; accession, protein reference found in NCBInr by LC−MS/MS; accession after BlastP, protein reference found in NCBInr or Uniprot; Protein name found in NCBInr or Uniprot; Predicted protein subcellular localizations found by using WoLF PSORT, Predotar, TargetP, YLoc and WegoLoc.

development, particularly associated with starch synthesis, starchy endosperm cell death, and stress/defense.

MATERIAL AND METHODS

Material Preparation

Triticum aestivum cv Récital was cultivated in normal wheat growing season (November−July) in a green house at INRA, Clermont-Ferrand, France. Plants were grown under natural soil conditions, fertilized (120 kg N-ha^{-1}), watered as usual and were protected against fungi.

Grains from the middle of the ears were tagged at the date of anthesis. Air temperature close to the ears was recorded and varied between 14.8 and 24.2 °C. Daily mean air temperature was calculated, and the sum of mean temperatures was used to follow the developmental stage in thermal time (°Cd). Wheat grains were harvested every 50 °Cd from anthesis (0 °Cd) to maturity (1006 °Cd) and were stored at −80 °C until dissection.

Dissection was carried out under laminar air flow using binocular for grains of 152−1006 °Cd, whereas whole grains were used for 0−93 °Cd, since they were very small and not fully differentiated. We separated starchy endosperm after removing embryo, brush part and peripheral layers, for study of metabolic proteins. For each developmental stage, the dissected endosperm was weighed and ground in liquid nitrogen using pestle and mortar. The resulting powder was stored at −80 °C until analysis.

Extraction of Alg Proteins

Albumins and globulins (alg) were extracted with low concentrated salt solution (Phosphate 10 mM, NaCl 10 mM, pH 7.8), which was supplemented with a cocktail of plant protease inhibitors (Sigma, St Louis, MO, USA), and was mixed continuously at 4 °C for 2 h.[21] After centrifugation (8000*g*, 20 min), the soluble proteins were precipitated with acetone at −20 °C. The pellet of extracted alg proteins was then washed several times with acetone before being dried at room temperature. Alg pellets were dissolved in extraction buffer 4% (w/v) CHAPS, 7 M urea, 2 M thiourea, 70 mM DTT, 1% (v/v) IPG buffer (pH 3−11 NL), and protease inhibitor cocktail, then the protein content was measured using the Bradford method.[22] Protein extracts were either directly used for IEF or were stored at −80 °C until electrophoresis.

Two-Dimensional Separation

Three biological extracts with two replicates per extract were used for analysis of each developmental stage. IEF was performed using the IPGPhor II apparatus (GE Healthcare, Uppsala, Sweden) on 24 cm Immobiline dry strips of 3−11 nonlinear pH gradient. Rehydration of strips was performed overnight at room temperature with 460 μL of solution containing 7 M urea, 2 M thiourea, 1% (v/v) IPG buffer (pH 3−11), 4% (w/v) CHAPS, 1.2% (v/v) destreak reagent (GE Healthcare, Uppsala, Sweden) and few grains of bromophenol blue. Protein extracts (300 μg) were cup-loaded on the acidic side of the strip and IEF was carried out by applying a cumulative voltage of 90 kVh. Following IEF, equilibration of strips, SDS-PAGE and gel staining using CBB G250, were performed as described earlier.[12]

Image and Statistical Analysis

G-800 (GE Healthcare, Uppsala, Sweden) scanner was used to obtain gel images that were then analyzed using SameSpots v4.1 (Nonlinear dynamics, Newcastle, UK). Proteins with fold change ≥1.8, *p*-value of ANOVA and with a *q*-value (measure of false positives in data) less than 0.05 were considered significant. Normalized volume values were used for statistical tests. Hierarchical cluster analysis (HCA) was computed using significant spots (950) for protein clustering according to Pearson's distance.

Protein Identification by LC−MS/MS

A total of 580 out of 950 (significantly varied protein spots), were excised from gels. The spots were destained and digested using the method described previously.[12] For LC−MS/MS analysis, peptide mixtures were analyzed by online nanoflow liquid chromatography using the Ultimate 3000 RSLC (Dionex, Voisins le Bretonneux, France) with nanocapillary columns of

15 cm length × 75 μm i.d (Acclaim Pep Map RSLC, Dionex). The solvent gradient increased linearly from 4 to 50% ACN in 0.5% formic acid at a flow rate of 300 nL/min for 30 min. The elute was then electrosprayed in a LTQ-VELOS mass spectrometer (Thermo Fisher Scientific, Courtaboeuf, France) through a nanoelectrospray ion source which was operated in a CID top 10 mode (i.e., 1 full scan MS and the 10 major peaks in the full scan were selected for MS/MS). Full-scan survey MS spectra were acquired with 1 microscan (m/z 400−1400). Dynamic exclusion was used with 2 repeat counts, 30 s repeat duration and 60 s exclusion duration. For MS/MS, isolation width for ion precursor was fixed at 3 m/z; fragmentation used 35% normalized collision energy at the default activation q of 0.25.

Thermo Proteome Discoverer v1.2 was used for raw data file processing. For protein identification, the NCBInr viridiplantae protein database was combined with sequences of human keratin contaminants. The following parameters were considered for the searches: peptide mass tolerance was set to 1.5 Da, fragment mass tolerance was set to 0.8 Da and maximum of two missed cleavages were allowed. Variable modifications were methionine oxidation (M) and carbamidomethylation (C) of cysteine. A protein was considered valid when a minimum of three unique peptides originating from one protein showed statistically significant ($p < 0.01$) Mascot scores (http://www.matrixscience.com/search_form_select.html). For several identification results of one protein spot, we selected the one with highest score for their functional classification (Table 1). Selection was based on the most appropriate taxonomy when Mascot reported alternatives with the same score. High confidence protein identifications in each spot are presented in Table S1 (Supporting Information). During grain development some protein identifications resulted in spots with unknown function. For these proteins, we made a BLASTP in NCBInr or UniProt using the protein sequences identified by LC−MS/MS.[23] The best hits were selected with at least 90% sequence similarity and significant score values.

Proteins were then classified according to KEGG PATHWAY database (http://www.genome.jp/kegg/pathway.html) and gene ontology. We established composite expression profiles for each functional category by summing normalized volumes of all the protein spots representative of that category for each developmental stage and then the mean value of 6 replicates was used to draw the expression curves.

Prediction of Subcellular localization

Subcellular localizations of identified proteins were predicted using the following programs: WoLF PSORT,[24] Predotar,[25] TargetP,[26] YLoc[27] and WegoLoc.[28] Subcellular localizations were added in Table 1, if at least three programs predicted the same localization of a protein at subcellular level.

■ RESULTS

Wheat endosperm alg proteins were studied at 21 stages during kernel development. Only one gluten protein (gamma-gliadin) was identified, which confirms the effectiveness of the protein separation method. A proteome map of metabolic proteins was developed for the whole kernel development. A total of 1780 spots were detected over the gels, among which 950 were significant on the basis of fold change and p- and q-values. The proteins above 80 kDa were abundant at early stages while low molecular weight proteins appeared progressively toward grain maturity. There was also a shift of proteins on 2D plane from acidic to basic zone, with maturity of grain. (Figure S1, Supporting Information)

In PCA results, the first component PC1 and second PC2 explained a total variance of 50.8% (32.5 and 18.3% respectively). Four major phases of development could be differentiated, i.e., (1) 0−93 °Cd; (2) 152−192 °Cd; (3) 247−764 °Cd; (4) 804−1006 °Cd. In addition, three subphases in phase 3 while two in phase 4 were identified (Figure S2, Supporting Information). The same grouping was found by plotting PC1 against PC3 (total variance 38.5%), and PC2 against PC3 (total variance 24.3%). A slight decrease in spot number was observed toward maturity (Figure 1a). The

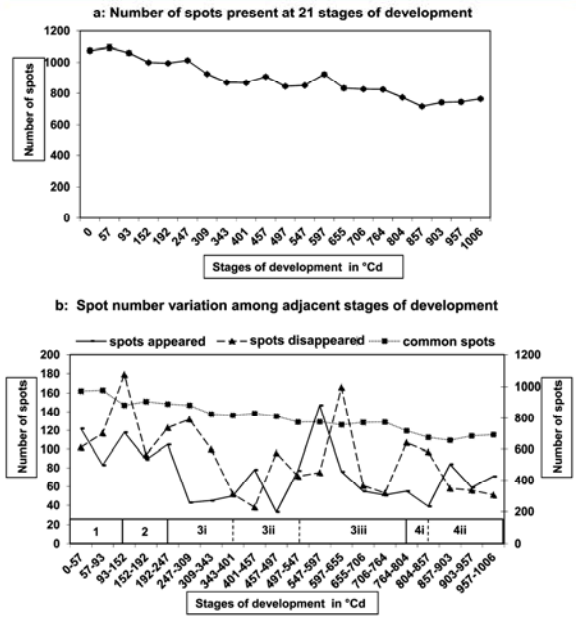


Figure 1. (a) Total spot number present at each stage of development (21 stages). (b) Number of spots that remain common and that appear and disappear between two adjacent stages. Stages of development are on the x-axis, while numbers of spots are represented on the y-axis.

number of spots that were common and that appeared and disappeared between two adjacent stages are presented in Figure 1b.

Protein Accumulation Clusters

Protein clustering revealed nine expression profiles (P1−P9) using HCA (Figure S3, Supporting Information). Expression tendency of protein spots in corresponding profiles is shown in Figure 2. On the basis of these curves, profiles were grouped in four main clusters as follows:

(i) Early accumulation cluster (EAC) grouped profiles 1 and 2, P1 with highest accumulation between 0 and 93 °Cd and P2 with a peak at 152 °Cd and then a rapid decrease. These are representative of proteins present in first and second phases of development (Figure S2, Supporting Information).

(ii) Early mid accumulation cluster (EMAC) included profiles 3 and 4, these two profiles peaked at 152 °Cd and 93 °Cd respectively and then decreased with grain maturity. In P3 proteins accumulated again at 903 °Cd

dx.doi.org/10.1021/pr201110d | J. Proteome Res. 2012, 11, 2754−2773

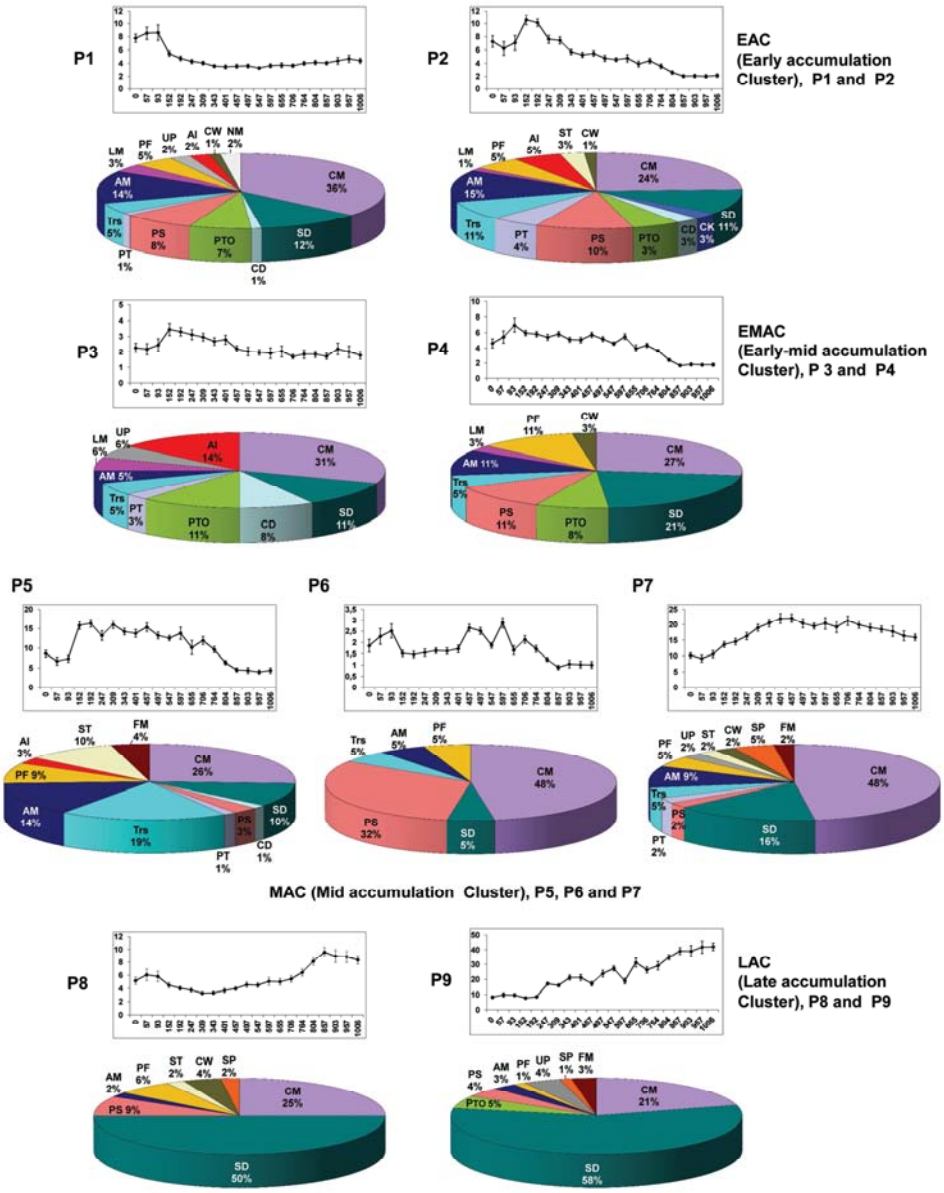


Figure 2. Functional distribution of proteins identified in each of the nine profiles. Each profile evolution is indicated by the curve above the corresponding pie chart, normalized volume is on the vertical axis, and stages of development (in °Cd) are on the horizontal axis with SD bars where $n = 6$. The percentage of identified spots included in each category is indicated. **AI**, ATP interconversion; **AM**, amino acid metabolism; **CD**, cell division; **CK**, cytoskeleton; **CM**, carbohydrate metabolism; **CW**, cell wall related; **FM**, folic acid metabolism; **LM**, lipid metabolism; **NM**, nucleotide metabolism; **PF**, protein folding; **PS**, protein synthesis; **PT**, protein transport; **PTO**, protein turn over; **SP**, storage proteins; **ST**, signal transduction; **SD**, stress/defense proteins; **Trs**, transcription/translation; **UP**, unknown proteins.

while in P4 from 804 °Cd their expression decreased significantly.

(iii) Mid accumulation cluster (MAC) was represented by P5, P6 and P7. In P5, proteins were mainly present between 152 and 706 °Cd while in P7 they accumulated between 401 and 706 °Cd. This cluster represented the proteins present largely during second and third phase of development. In contrast to P5 and P7, in P6 proteins were abundant at first and third phase of development with comparatively very low expression during second and fourth phase.

(iv) Late accumulation cluster (LAC) was comprised of P8 and P9. Proteins in P8 were accumulated particularly at later stages with a peak at 857 °Cd. In P9, a gradual increase in expression was observed from 247 °Cd which peaked at 1006 °Cd. Proteins of the fourth developmental phase were accumulated mainly in this cluster.

dx.doi.org/10.1021/pr201110d | J. Proteome Res. 2012, 11, 2754–2773

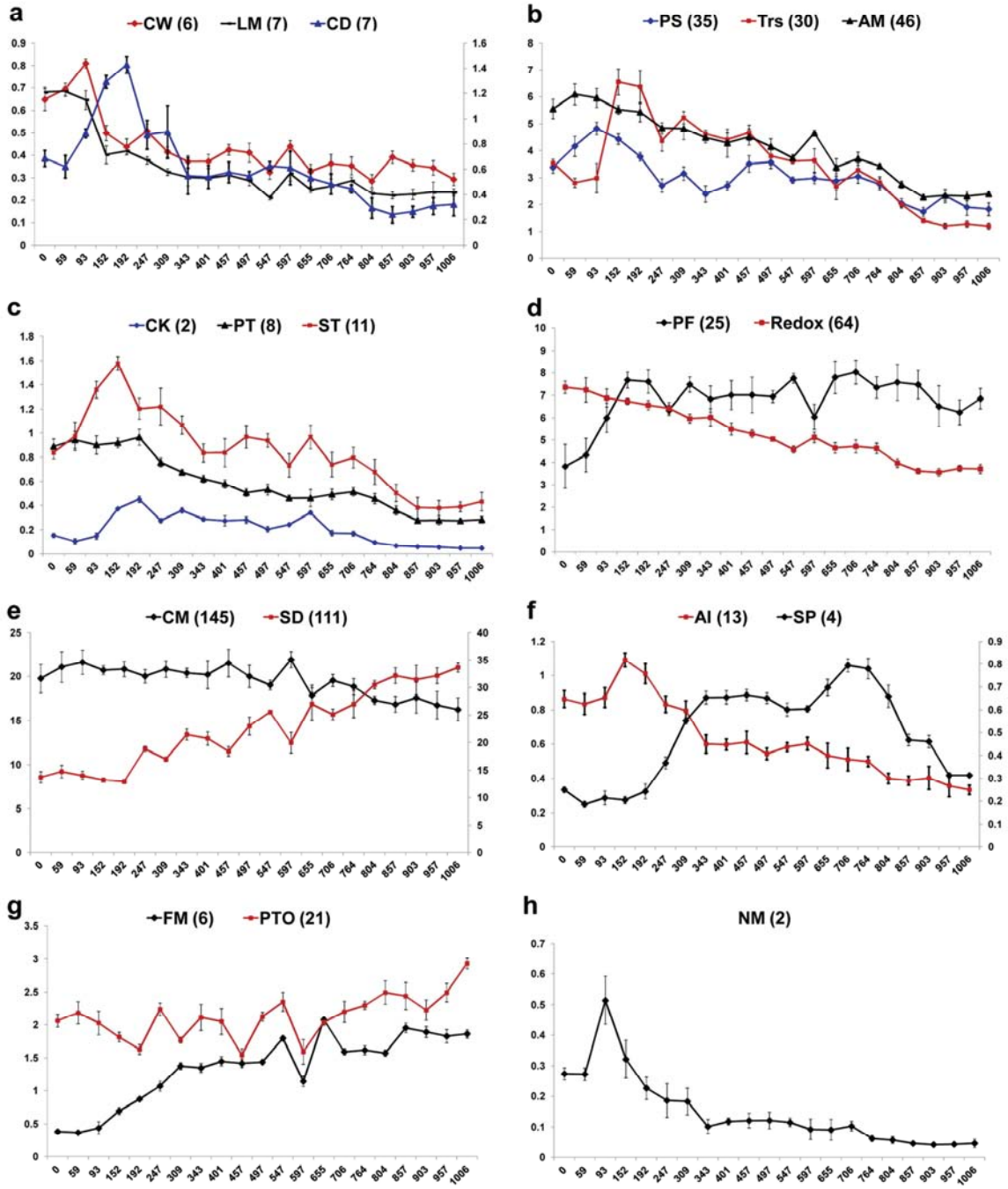


Figure 3. (a–h) Composite expression profiles of functional classes. The number of spots for the same category is indicated in parentheses. Each point is representative of the mean composite expression of normalized volume values of proteins (*y*-axis) at stages of development in °Cd (*x*-axis) and standard deviation $n = 6$.

Protein Functional Classification

Proteins were categorized in 17 different functional classes, using identified protein spots (487 out of 580). These spots were representative of all nine profiles (Table 1). This distribution differs not only between accumulation clusters but also within profiles of the same cluster (Figure 2).

The diversity of function was maximum in the early accumulation cluster (P1 and P2) with 14 different representative functions (Figure 2). In this cluster, dominating categories were carbohydrate metabolism (CM) with a 36% contribution in P1 and 24% in P2, followed by amino acid metabolism (AM) 14 and 15% respectively. Proteins related to

dx.doi.org/10.1021/pr201110d | *J. Proteome Res.* 2012, 11, 2754–2773

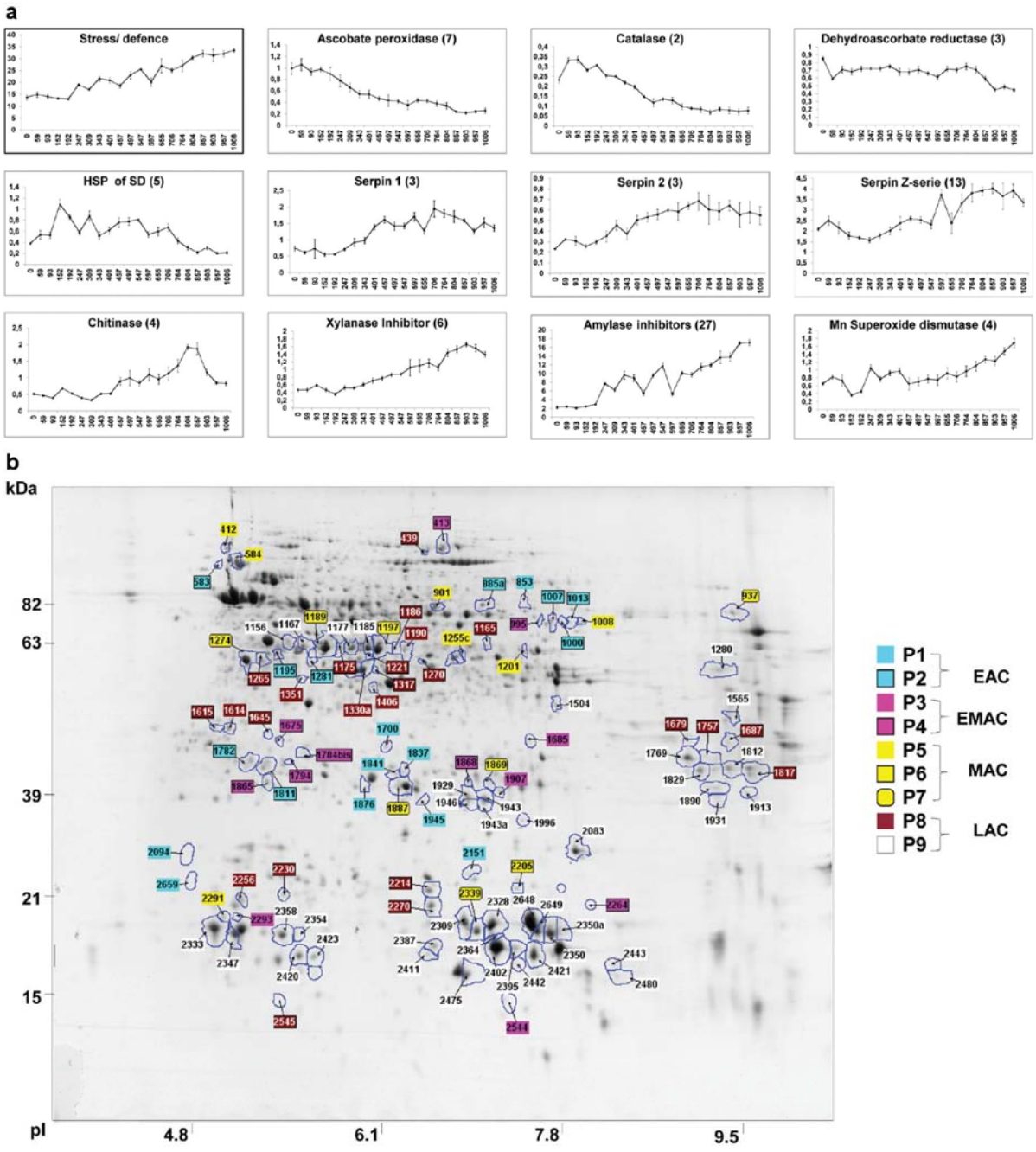


Figure 4. (a) Stress/defense and its subcategories representation, through their composite expression profiles. Composite expression of mean normalized volume values (*y*-axis) at 21 stages of development in °Cd (*x*-axis) with SD bars where $n = 6$. Total number of spots for each profile is indicated in parentheses. (b) Proteome map of stress/defense related proteins. Gel image is of 497 °Cd, and colored tags represent corresponding profiles of identified spot.

transcription/translation (Trs), ATP interconversion (AI), and protein transport (PT) were relatively more abundant in P2 than in P1.

P3 and P4, profiles of EMAC, were also dominated by CM (31 and 27% respectively). Significant changes in stress/defense (SD) (increase from 11 to 21%) AM (increase from 5 to 11%) and lipid metabolism (LM) (decrease from 6 to 3%) between P3 and P4 were observed. In this cluster, protein folding (PF), protein synthesis (PS), and cell wall (CW) related proteins were specific to P4 while cell division (CD) and AI were only identified in P3.

Similar to EAC and EMAC, CM proteins were the dominant category also in MAC and increased from 26 to 48%, whereas a considerable decrease in Trs (19 to 5%) and ST (10 to 2%) was

dx.doi.org/10.1021/pr201110d | J. Proteome Res. 2012, 11, 2754–2773

observed. Notably, P6 displayed minimum (only 6) diversity of functions. The majority were CM (48%) and PS (32%), while AM, SD, PF and Trs were present at 5% each. In P7, SD (16%) and AM (9%) followed the largest category of CM proteins.

Remarkably, in LAC, dominance of CM was replaced by SD related proteins (50% in P8 and 58% in P9), while CM was reduced to 25 and 21% in P8 and P9 respectively. PS, AM and ST were significantly reduced while LM, Trs, CD, CK, AI and PT vanished completely from P9.

Composite Expression Profiles of Protein Functional Classes

This composite expression was calculated by summing normalized spot volumes of a given functional category, for all corresponding proteins[14] (Figure 3).

Proteins involved in metabolism (LM, AM, CM, and NM) were abundant (Figure 3a,b,e,h) at early stages except folic acid metabolism (FM) (Figure 3g) where proteins accumulated toward grain maturity. CM as compared to AM, NM and LM showed nearly constant behavior throughout development with a slight gradual decrease in later developmental stages. Trs and ST attained a peak at 152 °Cd (Figure 3b,c) while PS and CD showed highest expression at 93 °Cd and 192 °Cd respectively (Figure 3b,a). CW related proteins were abundant at three early stages with a peak at 93 °Cd and afterward their level remained nearly constant (Figure 3a). Proteins related to AI category displayed a peak of expression at 152 °Cd, while those of PT were also abundant during early stages between 0 °Cd to 192 °Cd (Figure 3f,c).

Redox related proteins decreased with grain maturation (Figure 3d). At mid and mid late stages, proteins of CK, PF and SP were the dominating functional proteins (Figure 3c,d,f). Preponderance of SD and protein turn over (PTO) related proteins were shown at later stages of development (Figure 3e,g). SD protein subcategories are presented in Figure 4a. Although the SD category was dominant in later stages, composite expression profiles of its subcategories showed that all SD proteins were not dominant in LAC, but rather some were also abundantly present in EAC and MAC.

Prediction of Subcellular Protein Localization

Subcellular localization of identified proteins varied during development. In EAC proteins were scattered in eight different locations as compared to five in LAC. Cytoskeleton (cysk) proteins were found only in P1 and P2, nuclear (nucl) proteins from P1 to P7, while those of peroxisome (perox) in P1, P4 and P5. Proteins in cytosol (Cyto), endoplasmic reticulum (ER), extracellular region (extr), plastid (P), and mitochondria (M) were present throughout development. However, in Cyto they were abundant during early stages, in M and P during early and mid stages while those of ER and Extr were dominant in later stages (Figure S4, Supporting Information). Pathway and functional classifications were assigned for SD category proteins on the basis of ontology, and on metabolic features particularly established in plants (http://metacyc.org) (Figure S5, Supporting Information).[29,30]

■ DISCUSSION

Wheat endosperm is a triploid tissue in which 80−85% of extractable proteins are gliadins and glutenins. Alg are less abundant proteins and consequently have been given less attention for grain quality improvement studies. Hand dissection of endosperm followed by protein extraction methods specific for alg proved to be efficient, as also revealed by previous studies to determine their 2D patterns. In the present study, 21 stages were selected to develop a comprehensive proteome map of alg proteins during kernel development.

The total number of spots during development remained nearly constant with a slight decrease toward grain maturity. However, between developmental stages, relative abundance of individual spots was different, since image analysis detected 950 spots varying significantly between two or more stages. This protein spot variation provides evidence for coregulation, probable gene function that is activated or inactivated at specific phases and relationships between phases.

Nine accumulation profiles were distinguished by HCA and were grouped in four main clusters during alg temporal analysis. Proteins involved in carbohydrate metabolism, amino acid metabolism, and stress/defense were identified in all expression profiles. Transcription/translation was not identified in P8 and P9, while protein synthesis and protein folding mechanisms were not found in P3. Few functional categories appeared only in one or two accumulation profiles.

Early Accumulation Clusters (EAC and EMAC)

In early accumulation clusters (EAC and EMAC), EAC represented maximum functional diversity, with 14 different functions in both P1 and P2 (Figure 2). The highest number of proteins was identified in P1 among nine profiles, with a majority of proteins involved in CM. In early stages of development, wheat caryopsis changes dramatically in size and shape, and study of caryopsis structure revealed that, in endosperm at 3 DAP (≈ 60 °Cd), a single layer of free nuclei is present only around the central vacuole.[31] At 6 DAP (≈110 °Cd), in parallel to caryopsis development, these nuclei are present in the whole central region that was occupied by the central vacuole because of cellularization of peripheral endosperm. All major cell types are differentiated around 13 DAP (≈ 250 °Cd).[32] CD and CK related proteins are thought to play significant roles in these developmental processes. Microtubules (CD) and actin filaments (CK) are not only important for cell growth and division but they also target organelles and vesicles.[33] To participate in above-mentioned tasks, CD and CK were maximum at 192 °Cd, the period before the start of grain filling (Figure 3a,c). Enzymes for LM were also abundant in EAC and EMAC. These enzymes, together with CD and CK related proteins, take part in cell division, cell expansion, differentiation, sequential and continuous processes involving cessation of sucrose uptake, and accumulation of storage products.

Sucrose and nitrogen are taken up by seeds in the form of amino acids from the apoplast,[34] and generally biosynthesis of seed storage protein is dependent on enzymes of nitrogen metabolism.[35,36] Enzymes for synthesis of amino acids such as phenylalanine, tyrosine, serine, homocysteine, cysteine, proline, glycine, aspartate, glutamic acid, methionine, threonine were identified. In addition, two other amino acids, acetohydroxyacid synthase (involved in biosynthesis of leucine, isoleucine and valine)[37] and ketolacid reductoisomerase responsible for isoleucine and valine biosynthesis were also found. These enzymes were abundant in EAC and EMAC (prestorage phase) and attenuated with grain maturity.

Mid Accumulation Cluster (MAC)

In this cluster we found some proteins of prestorage and mainly of storage phase. This cluster was dominated by CM, SD, Trs, and AM. CM continued to be the principal metabolism, and in

P6 and P7 its participation was highest (48%) among all profiles (Figure 2). A study by Lunn et al. supported the proposal that the changes induced in the rate of starch synthesis by sucrose are mediated by trehalose-6-phosphate (Tre 6P), as they found it as a signaling metabolite of sugar status in plants.[38] In wheat, Tre 6P expression is tissue and developmental stage dependent, at 7 DAP (≈130 °Cd), its level was nearly same in both maternal and filial tissues but at grain filling stage 17 DAP (≈ 320 °Cd), accumulation was mostly restricted to endosperm.[39] Enzymes involved in sucrose and starch metabolism were found abundantly in this cluster, complementing the above results where starch synthesis was higher at 17 DAP (≈ 320 °Cd).

Starch biosynthesis and accumulation is the major process of grain filling and four principal enzymes sucrose synthase, ADP glucose pyrophosphate, starch branching enzyme (SBE) and starch synthase are involved.[40] Accordingly, sucrose synthase, SBE and ADP glucose pyrophosphate relative abundances were found maximum between 152 and 706 °Cd with peak at 597 °Cd. A starch synthesis regulatory enzyme, alpha 1,4-glucan phosphorylase,[41] accumulated also between 152 and 706 °Cd. These four curves indicated that, for accumulation of enzymes involved in the biosynthesis of starch, the critical period lies between 152 and 706 °Cd. (Figure S6a, Supporting Information)

Programmed cell death (PCD) in cereal endosperm moves in parallel to grain filling. Studies on developing wheat endosperm revealed that PCD was first detected at 16 DAP (≈ 300 °Cd), and finished with starch and protein accumulation.[42,43] In our study, the four identified SP peaked at 706 °Cd, (Figure 3f) the period of endosperm development completion.[43] They had an expression curve similar to those of MAC, i.e., no or very low expression in the early and late stages and high expression in the middle stages of development. At the start of grain filling a high sugar level may be involved indirectly for ethylene production, which is an important key for PCD onset in developing endosperm.[44] At 30 DAP (≈ 580 °Cd), PCD penetrates in whole endosperm, and its products (mainly proteases in plants) increased as revealed by the peak of proteasomes, which increased after 597 °Cd until 1006 °Cd (Figure S6b, Supporting Information). A reduction in the accumulation of translation initiation factors decreased protein synthesis which is probably also a result of PCD (Figure 3b). Cell death is also strongly influenced by oxidative stress.

The absence of stomata, as well as the presence of several outer seed layers, acts as diffusion barriers within tissues and results in hypoxia that leads to low energy production. The redox mechanism was highly active at early stages of development because of the permeability of the young pericarp and photosynthetic mechanism of peripheral grain layers that are still green.[14] In developing seeds, a struggle for energy and oxygen was noticed from the reserve accumulation stage to desiccation period,[45] and there we observed a decrease of the redox curve (Figure 3d). During wheat kernel development, ambient oxygen was found to be hypoxic.[46] With less oxygen availability, primary metabolism activity slows down, or switches to metabolic pathways with less ATP consumption.[47,48] Accordingly, AI and CW in addition to nearly all metabolic processes (CM, AM, LM, NM), except FM, decreased (Figure 3). These metabolic changes were found to be due to membrane alterations and less seed respiration caused by an adaptive response to avoid or to postpone plant tissues from suffering anoxia and its concomitant negative effects.[49]

Late Accumulation Cluster (LAC)

P8 and P9 represent the proteins of LAC that were found to be abundant in grain maturity stages. Composite protein expression curves revealed that these stages were dominated by the proteins involved in SD, PTO and FM. An increase in PTO curve was mainly due to proteasome proteins discussed above in PCD. In P8 one gamma-gliadin was identified. Normally gamma-gliadin possesses eight cyteins. Sequence inquiry revealed strong similarity with gliadin/avenin-like seed proteins (commonly found in alg preparations) and an even number (six) of cystein giving three intramolecular disulfide bonds. This made gamma-gliadin partially soluble and unpolymerized with other storage proteins; hence it was extracted with soluble proteins of alg.

SD Enzymes

In SD, proteins related to both biotic and abiotic stresses were found abundantly in LAC, with continuous increase from 247 °Cd up to 1006 °Cd. These proteins at EAC and EMAC were present from acidic to central (neutral) part of the gels, while in the basic zone, there were only the proteins of later stages particularly of P8 and P9 (Figure 4b).

Subcellular localization of SD proteins revealed that the majority of these proteins were in cytosol (35%) and extracellular region (29%); they were comparatively less abundant in P (20%) and ER (11%) while only few were observed in perox (3%), M (2%) and nucl (1%). In early developmental stages proteins were abundant mostly in cytosol, with only little number in other localizations. In parallel with grain development they appeared abundantly in several localizations, as in P9 they were predicted significantly in four different localizations, i.e., cyto, P, Extr, and ER. (Figure 5).

Metabolic pathways of SD category enzymes are presented in (Figure S5, Supporting Information). This provided a rapid overview of SD enzymes and also specifies their position in the developmental scale. ROS related enzymes were found to be abundant during EAC and EMAC, as can be seen from composite expression curves of catalase, ascorbate peroxidase,

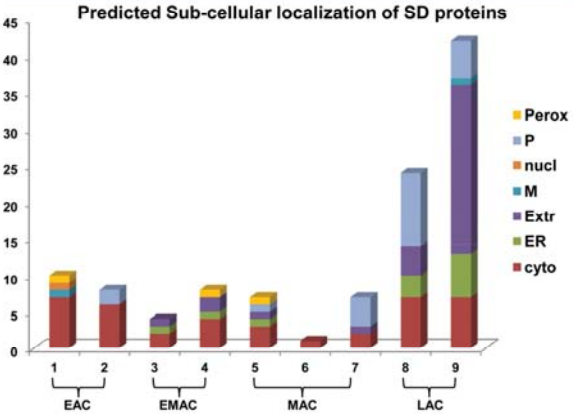


Figure 5. Predicted stress/defense protein distribution in different subcellular localizations. *X*-axis represents profiles (P1−P9), while *y*-axis represents the number of identified proteins in corresponding profiles. **Cyto**, cytosol; **ER**, endoplasmic reticulum; **Extr**, extra-cellular region; **M**, mitochondria; **nucl**, nuclear; **P**, plastid; **Perox**, peroxisome.

dx.doi.org/10.1021/pr201110d | J. Proteome Res. 2012, 11, 2754−2773

dehydro ascorbate reductase (Figure 4a). These enzymes were abundant in cytosol in early stages for processes such as detoxification and dehydrogenation. In these stages enzymes present in Perox were for catabolism of long chain fatty acids, and LM abundance thus decreased toward maturity (Figure 3a). Abiotic stress related proteins such as salt tolerant protein and Wali 7 (Al stress related) were observed also in early accumulation clusters, probably for assistance of healthy seed development.

Five heat shock proteins involved in SD were observed in P2, P4, P5 and P8. High molecular weight HSPs such as 101 were found in cyto; these are chaperons in nature and are involved in refolding of proteins that are denatured during EMAC (spot 413) and LAC (spot 439). In MAC, HSP90 were located in cyto (584) and ER (412). They were active in parallel with PF (Figure 3d) as they help in stabilizing proteins prior to complete folding. Briefly, HSP network was found functional throughout development during three different compartments.

Inhibitors of xylanase, alpha-amylase and serine proteinase along with chitinases were found to be abundant in later stages (Figure 4a), and their possible role in different pathways is indicated (Figure S5, Supporting Information). In addition to their role in starch mobilization, alpha amylase inhibitors are involved in many other functions; for example, they serve as storage proteins, provide essential amino acids for human nutrition, have protective role and are also involved in wheat allergies.[50]

LAC proteins were predicted mainly in four different localizations. Serpins, which are a complex-natured protein family, were found in P, abundantly in MAC (P5 and P7) and in both profiles of LAC. It was found that wheat serpins may have a protective role for prolamins,[51] which are present around the amyloplasts. They probably also protect starch, which shared the same subcellular localization as serpins. Extracellular region proteins, which are either synthesized directly in this region or came through secretory pathway,[52] were principally inhibitors of alpha amylase and xylanase. These proteins while residing in plasma membrane help in protecting the cellular contents of dehydrated grain. Well-established grain protective mechanisms were found to be active during grain development, particularly in extracellular region at later stages.

Study of grain development with monitoring at every 50 °Cd provided not only an insight into the major events but also a more precise understanding of the occurrence of these events (in terms of composite expression profiles for a given functional category). In a previous study while comparing two stages, 10 DAP (≈200 °Cd) and 36 (≈700 °Cd), PT and AI proteins were found to be abundant at 700 °Cd.[19] In contrast we found that PT and AI peaked at 192 and 152 °Cd respectively, and then decreased with grain maturity (Figure 3c,f). This difference could result from the lower number of proteins identified (4 vs 8 for PT and 2 vs 13 for AI) in earlier study, which also revealed that SP expression peaked at 700 °Cd. The findings in the present experiment not only validated SP results, but also it was noticed that, after 700 °Cd, expression decreased significantly for SP (Figure 3f). This decrease of SP may be due to their remobilization for accumulation of other proteins in later stages. Seven PTO identified by Vensel et al. were found abundant at 200 °Cd, whereas no significant differences were found between 200 and 700 °Cd for 21 PTO identified in the present study. In addition, after 700 °Cd they increased toward grain maturity (Figure 3g).

Despite these differences, this study helped us to reveal the expression behavior throughout development for proteins, whose expression was found common between these two studies at both developmental stages. As for proteins that were observed abundantly at 200 °Cd, we additionally found that CM and AM decrease linearly, CD, ST and LM were highly abundant in early grain development stages and then decreased significantly, while CK proteins appeared uniformly between 152 and 597 °Cd. Similarly, the SD category, which increased towards grain maturity in both studies, showed a linear increase from 247 °Cd until maturity.

This study with 21 stages unraveled the alg protein expression during whole grain development and provided valuable information that complemented the events known up to present. In addition to four main phases of development, three subphases in phase three and two in phase 4 were evident. In this study we constructed an alg proteome reference map with nine submaps on the basis of corresponding expression profiles not so far described, as previous studies were on limited scale of development in terms of number and interval.[18−20] Additional areas need to be explored, such as comparison between expression profiles of metabolic processes by mining of different wheat tissues or by performing the different "omics" analysis for development of comprehensive understanding of wheat grain development. This reference study could be used for investigation of biotic and abiotic stress responses. Study of alg and gluten proteins during development would help to explore the influence of soluble proteins on accumulation of storage proteins.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S6 and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: branlard@clermont.inra.fr. Tel.: +33 (0)473624316. Fax: +33 (0)473624453.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge gratefully Didier Viala for MS analysis, Saïd Mouzeyar, Emmanuelle Bancel, and Felicity Vear for critical reading, and Higher Education Commission, Pakistan for financial support.

■ ABBREVIATIONS:

°Cd, degree days after pollination; alg, albumins globulins; EAC, early accumulation cluster; EMAC, early mid accumulation cluster; MAC, mid accumulation cluster; LAC, late accumulation cluster; HCA, hierarchical cluster analysis; PCA, principal component analysis

■ REFERENCES

(1) Bietz, J. A.; Wall, J. S. Wheat gluten subunits: Molecular weight determined by sodium deodecyl sulfate-polyacrylamide gel electrophoresis. *Cereal Chem.* **1972**, *49* (4), 416−430.

(2) Brown, J. W. S.; Flavell, R. B. Fractionation of wheat gliadin and glutenin subunits by two-dimensional electrophoresis and the role of

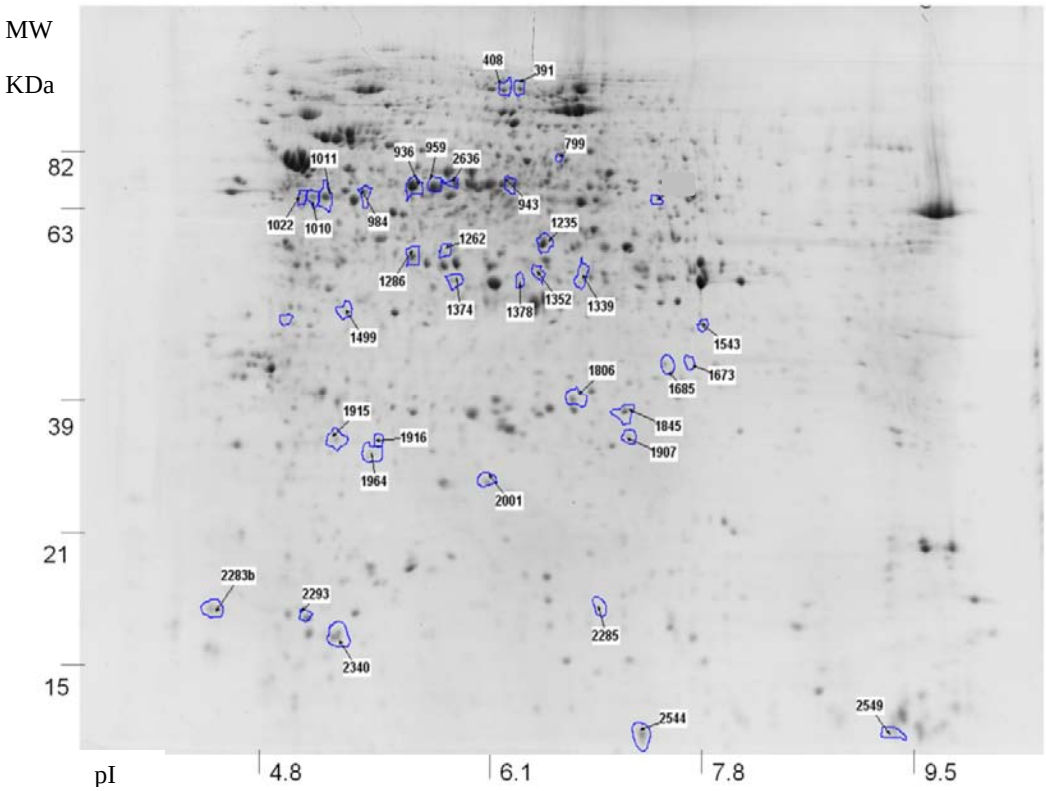
group-6 and group-2 chromosomes in gliadin synthesis. *Theor. Appl. Genet.* **1981**, *59* (6), 349−359.
(3) Islam, N.; Woo, S. H.; Tsujimoto, H.; Kawasaki, H.; Hirano, H. Proteome approaches to characterize seed storage proteins related to ditelocentric chromosomes in common wheat (*Triticum aestivum* L.). *Proteomics* **2002**, *2* (9), 1146−1155.
(4) Branlard, G.; Dumur, J.; Bancel, E.; Merlino, M.; Dardevet, M. Proteomic analysis of wheat storage proteins: A promising approach to understand the genetic and molecular bases of gluten components, *The Gluten proteins*, Royal Society of Chemistry: Cambridge, U.K., 2004, Issue 295, pp 30−33.
(5) Tosi, P.; Gritsch, C. S.; He, J. B.; Shewry, P. R. Distribution of gluten proteins in bread wheat (*Triticum aestivum*) grain. *Ann. Bot.* **2011**, *108* (1), 23−35.
(6) Pomeranz, Y. *Wheat: Chemistry and Technology, AACC Monograph Series*; American Association of Cereal Chemists: St Paul, MN, USA, 1988; Vol. *I*.
(7) Hill, K.; Horvath-Szanics, E.; Hajos, G.; Kiss, E. Surface and interfacial properties of water-soluble wheat proteins. *Colloids Surf., A* **2008**, *319* (1−3), 180−187.
(8) Singh, J.; Skerritt, J. H. Chromosomal control of albumins and globulins in wheat grain assessed using different fractionation procedures. *J. Cereal Sci.* **2001**, *33* (2), 163−181.
(9) Koch, K. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr. Opin. Plant Biol.* **2004**, *7* (3), 235−246.
(10) Wobus, U.; Weber, H. Sugars as signal molecules in plant seed development. *Biol. Chem.* **1999**, *380* (7−8), 937−944.
(11) Borisjuk, L.; Rolletschek, H.; Radchuk, R.; Weschke, W.; Wobus, U.; Weber, H. Seed development and differentiation: A role for metabolic regulation. *Plant Biol.* **2004**, *6* (4), 375−386.
(12) Tasleem-Tahir, A.; Nadaud, I.; Girousse, C.; Martre, P.; Marion, D.; Branlard, G. Proteomic analysis of peripheral layers during wheat (*Triticum aestivum* L.) grain development. *Proteomics* **2011**, *11* (3), 371−379.
(13) Finnie, C.; Melchior, S.; Roepstorff, P.; Svensson, B. Proteome analysis of grain filling and seed maturation in barley. *Plant Physiol.* **2002**, *129* (3), 1308−1319.
(14) Mechin, V.; Thevenot, C.; Le Guilloux, M.; Prioul, J. L.; Damerval, C. Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. *Plant Physiol.* **2007**, *143* (3), 1203−1219.
(15) Gallardo, K.; Le Signor, C.; Vandekerckhove, J.; Thompson, R. D.; Burstin, J. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* **2003**, *133* (2), 664−682.
(16) Houston, N. L.; Hajduch, M.; Thelen, J. J. Quantitative proteomics of seed filling in castor: comparison with soybean and rapeseed reveals differences between photosynthetic and non-photosynthetic seed metabolism. *Plant Physiol.* **2009**, *151* (2), 857−868.
(17) Singh, J.; Blundell, M.; Tanner, G.; Skerritt, J. H. Albumin and globulin proteins of wheat flour: Immunological and N-terminal sequence characterisation. *J. Cereal Sci.* **2001**, *34* (1), 85−103.
(18) Wong, J. H.; Cal, N.; Balmer, Y.; Tanaka, C. K.; Vensel, W. H.; Hurkman, W. J.; Buchanan, B. B. Thioredoxin targets of developing wheat seeds identified by complementary proteomic approaches. *Phytochemistry* **2004**, *65* (11), 1629−1640.
(19) Vensel, W. H.; Tanaka, C. K.; Cai, N.; Wong, J. H.; Buchanan, B. B.; Hurkman, W. J. Developmental changes in the metabolic protein profiles of wheat endosperm. *Proteomics* **2005**, *5* (6), 1594−1611.
(20) Gao, L. Y.; Wang, A. L.; Li, X. H.; Dong, K.; Wang, K.; Appels, R.; Ma, W. J.; Yan, Y. M. Wheat quality related differential expressions of albumins and globulins revealed by two-dimensional difference gel electrophoresis (2-D DIGE). *J. Proteomics* **2009**, *73* (2), 279−296.
(21) Debiton, C.; Merlino, M.; Chambon, C.; Bancel, E.; Decourteix, M.; Planchot, V.; Branlard, G. Analyses of albumins, globulins and amphiphilic proteins by proteomic approach give new insights on waxy wheat starch metabolism. *J. Cereal Sci.* **2011**, *53* (2), 160−169.
(22) Bradford, M. M. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal. Biochem.* **1976**, *72* (1−2), 248−254.
(23) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J. H.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389−3402.
(24) Horton, P.; Park, K. J.; Obayashi, T.; Fujita, N.Harada, H.Adams-Collier, C. J.. Nakai, K. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*; 2007, *35*, W585−W587.
(25) Small, I.; Peeters, N.; Legeai, F.; Lurin, C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **2004**, *4* (6), 1581−1590.
(26) Emanuelsson, O.; Nielsen, H.; Brunak, S.; von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **2000**, *300* (4), 1005−1016.
(27) Briesemeister, S.; Rahnenfuhrer, J.; Kohlbacher, O. YLoc-an interpretable web server for predicting subcellular localization. *Nucleic Acids Res.* **2010**, *38*, W497−W502.
(28) Chi, S. M.; Nam, D. WegoLoc: accurate prediction of protein subcellular localization using weighted Gene Ontology terms. *Bioinformatics* **2012**, DOI: 10.1093/bioinformatics/bts062.
(29) Buchanan, B. B.; Gruissem, W.; Jones, R. L. *Biochemistry and Molecular Biology of Plants*; American Society of Plant Physiology: Rockville, MD, 2000.
(30) Catusse, J.; Strub, J. M.; Job, C.; Van Dorsselaer, A.; Job, D. Proteome-wide characterization of sugarbeet seed vigor and its tissue specific expression. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (29), 10262−10267.
(31) Sabelli, P. A.; Larkins, B. A. The Development of Endosperm in Grasses. *Plant Physiol.* **2009**, *149* (1), 14−26.
(32) Drea, S.; Leader, D. J.; Arnold, B. C.; Shaw, P.; Dolan, L.; Doonan, J. H. Systematic spatial analysis of gene expression during wheat caryopsis development. *Plant Cell* **2005**, *17* (8), 2172−2185.
(33) Mayer, U.; Jurgens, G. Microtubule cytoskeleton: a track record. *Curr. Opin. Plant Biol.* **2002**, *5* (6), 494−501.
(34) Weber, H. Molecular physiology of legume seed development. *Ann. Rev. Plant Biol.* **2005**, *56*, 253−279.
(35) Balconi, C.; Rizzi, E.; Manzocchi, L.; Soave, C.; Motto, M. Analysis of in vivo and in vitro grown endosperms of high and low protein strains of maize. *Plant Sci.* **1991**, *73* (1), 1−9.
(36) Hernandez-Sebastia, C.; Marsolais, F.; Saravitz, C.; Israel, D.; Dewey, R. E.; Huber, S. C. Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low- and high-protein soybean lines. *J. Exp. Bot.* **2005**, *56* (417), 1951−1963.
(37) Duggleby, R. G.; Pang, S. S. Acetohydroxyacid synthase. *J. Biochem. Mol. Biol.* **2000**, *33* (1), 1−36.
(38) Lunn, J. E.; Feil, R.; Hendriks, J. H. M.; Gibon, Y.; Morcuende, R.; Osuna, D.; Scheible, W. R.; Carillo, P.; Hajirezaei, M. R.; Stitt, M. Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in Arabidopsis thaliana. *Biochem. J.* **2006**, *397*, 139−148.
(39) Martinez-Barajas, E.; Delatte, T.; Schluepmann, H.; de Jong, G. J.; Somsen, G. W.; Nunes, C.; Primavesi, L. F.; Coello, P.; Mitchell, R. A. C.; Paul, M. J. Wheat grain development is characterized by remarkable trehalose 6-phosphate accumulation pregrain filling: Tissue distribution and relationship to SNF1-related protein kinase1 activity. *Plant Physiol.* **2011**, *156* (1), 373−381.
(40) Yang, J. C.; Zhang, J. H.; Wang, Z. Q.; Xu, G. W.; Zhu, Q. S. Activities of key enzymes in sucrose-to-starch conversion in wheat grains subjected to water deficit during grain filling. *Plant Physiol.* **2004**, *135* (3), 1621−1629.
(41) Tickle, P.; Burrell, M. M.; Coates, S. A.; Emes, M. J.; Tetlow, I. J.; Bowsher, C. G. Characterization of plastidial starch phosphorylase in *Triticum aestivum* L. endosperm. *J. Plant Physiol.* **2009**, *166* (14), 1465−1478.

(42) Young, T. E.; Gallie, D. R. Analysis of programmed cell death in wheat endosperm reveals differences in endosperm development between cereals. *Plant Mol. Biol.* **1999**, *39* (5), 915−926.
(43) Rui, L.; Sheng-Yin, L.; Zhen-Xiu, X. Programmed cell death in wheat during starchy endosperm development. *J. Plant Physiol. Mol. Biol.* **2004**, *30* (2), 183−188.
(44) Young, T. E.; Gallie, D. R. Programmed cell death during endosperm development. *Plant Mol. Biol.* **2000**, *44* (3), 283−301.
(45) Angelovici, R.; Galili, G.; Fernie, A. R.; Fait, A. Seed desiccation: a bridge between maturation and germination. *Trends Plant Sci.* **2010**, *15* (4), 211−218.
(46) van Dongen, J. T.; Roeb, G. W.; Dautzenberg, M.; Froehlich, A.; Vigeolas, H.; Minchin, P. E. H.; Geigenberger, P. Phloem import and storage metabolism are highly coordinated by the low oxygen concentrations within developing wheat seeds. *Plant Physiol.* **2004**, *135* (3), 1809−1821.
(47) Huang, S. B.; Colmer, T. D.; Millar, A. H. Does anoxia tolerance involve altering the energy currency towards PPi? *Trends Plant Sci.* **2008**, *13* (5), 221−227.
(48) Vigeolas, H.; van Dongen, J. T.; Waldeck, P.; Huhn, D.; Geigenberger, P. Lipid storage metabolism is limited by the prevailing low oxygen concentrations oilseed rape. *Plant Physiol.* **2003**, *133* (4), 2048−2060.
(49) Geigenberger, P. Response of plant metabolism to too little oxygen. *Curr. Opin. Plant Biol.* **2003**, *6* (3), 247−256.
(50) Altenbach, S. B.; Vensel, W. H.; Dupont, F. M. The spectrum of low molecular weight alpha-amylase/protease inhibitor genes expressed in the US bread wheat cultivar Butte 86. *BMC Res. Notes* **2011**, *4*, 242.
(51) Ostergaard, H.; Rasmussen, S. K.; Roberts, T. H.; Hejgaard, J. Inhibitory serpins from wheat grain with reactive centers resembling glutamine-rich repeats of prolamin storage proteins — Cloning and characterization of five major molecular forms. *J. Biol. Chem.* **2000**, *275* (43), 33272−33279.
(52) Rose, J. K. C.; Lee, S. J. Straying off the Highway: Trafficking of secreted plant proteins and complexity in the plant cell wall proteome. *Plant Physiol.* **2010**, *153* (2), 433−436.

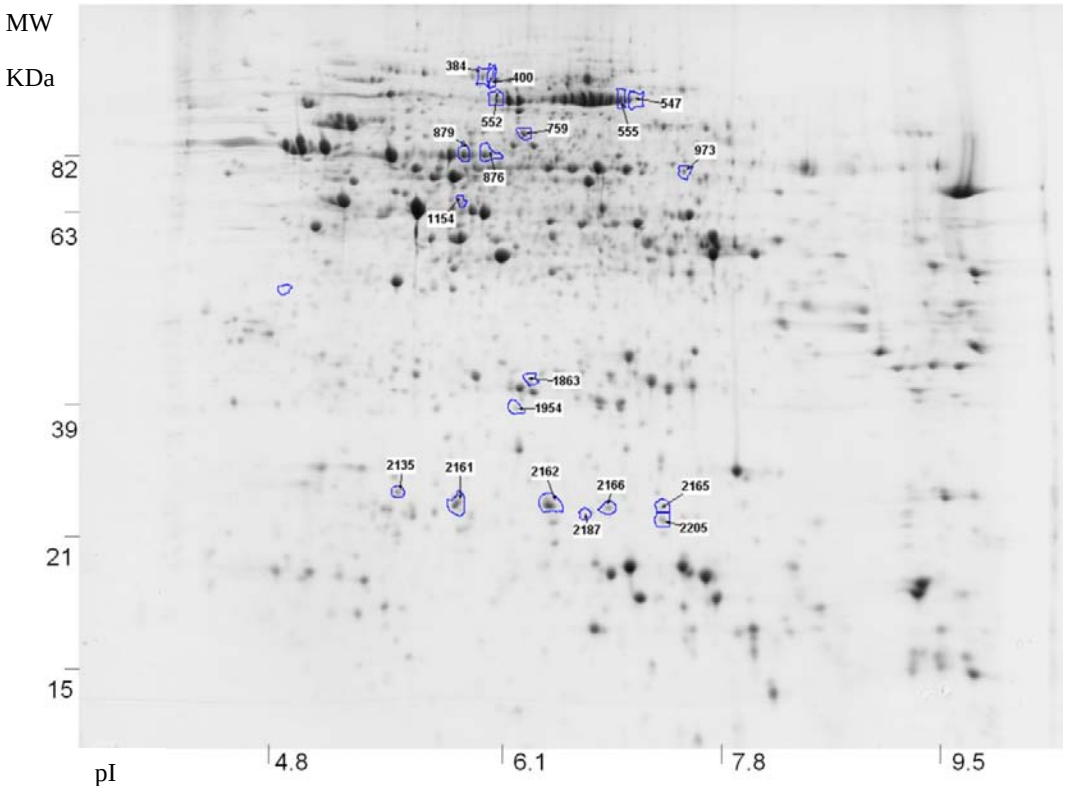
Additional Data

Supplemental Figure S1: Proteins separated by 2-DE, numbered spots were identified (Supplemental Table S1). a) 152°Cd; b) 457°Cd; c) 903°Cd. Spots abundance shifts gradually from upper to lower part and from acidic to basic zone of gel, with kernel development.

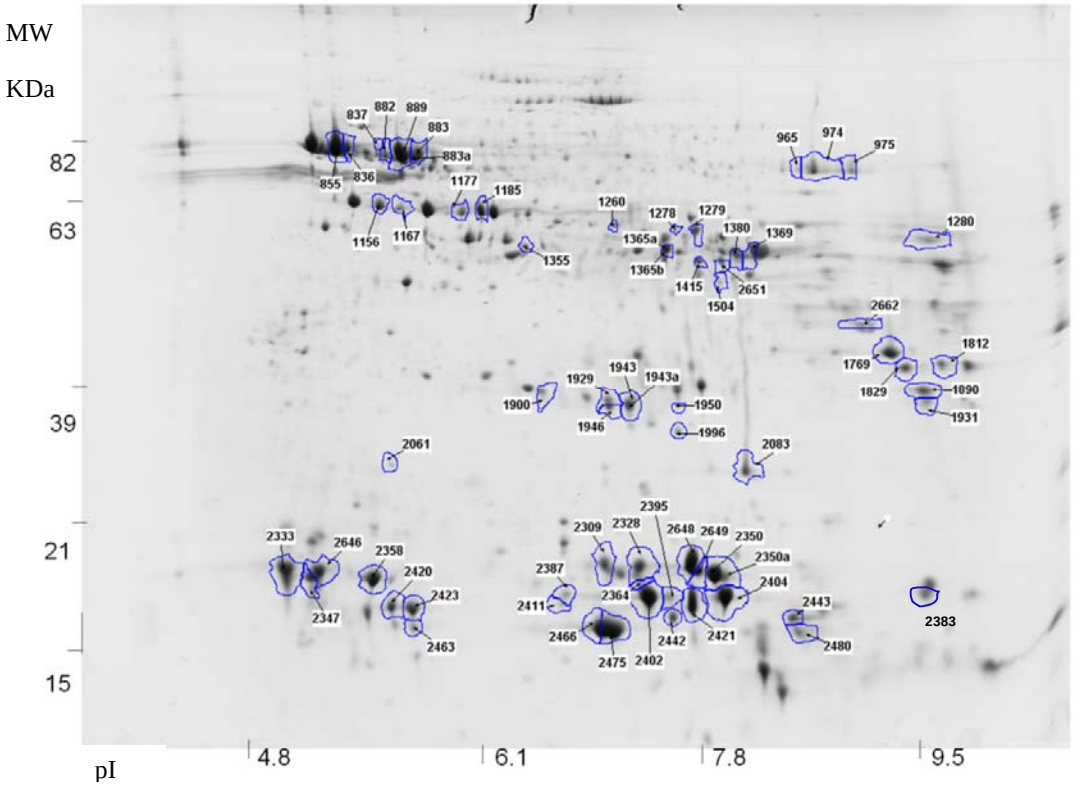
a) Gel of stage 152°Cd (identified spots of P3)



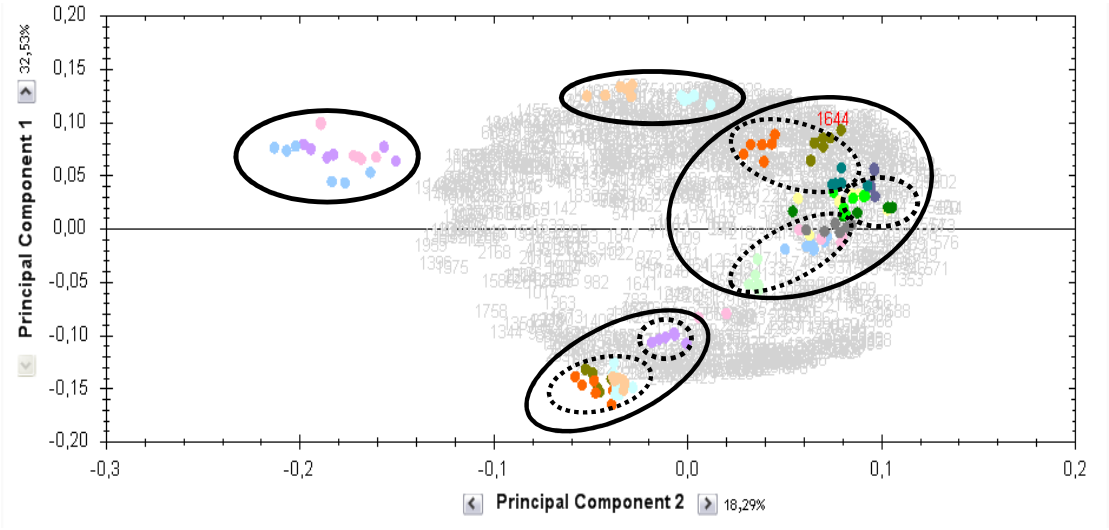
b) Gel of stage 457°Cd (identified spots of P6)

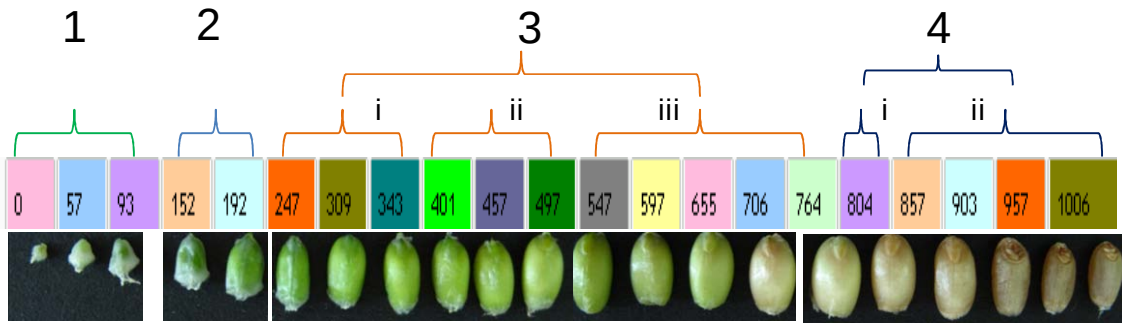


c) Gel of stage 903°Cd (identified spots of P9)

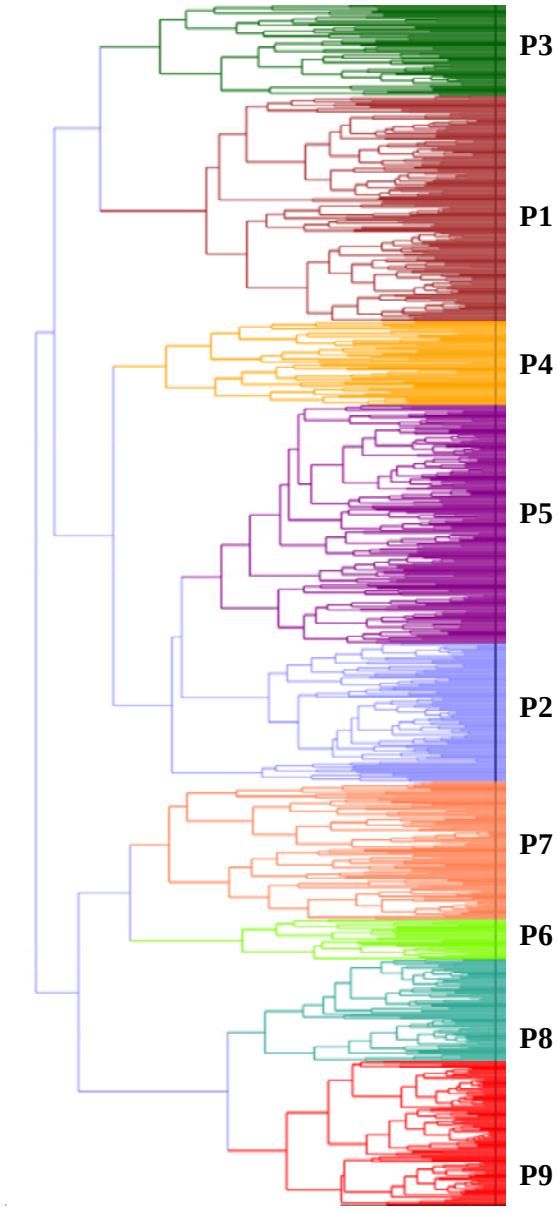


Supplemental Figure S2: Principal component analysis for 21 x 6 gels. PC1=32.5%, PC2=18.3%, PC3=6.0%. Four main phases of development were identified, with three sub-phases in phase 3 and two sub-phases in phase 4. Four main phases of development could be observed: 1) 0-93°Cd; 2) 152-192°Cd; 3) 247-764°Cd; 4) 804-1006°Cd.

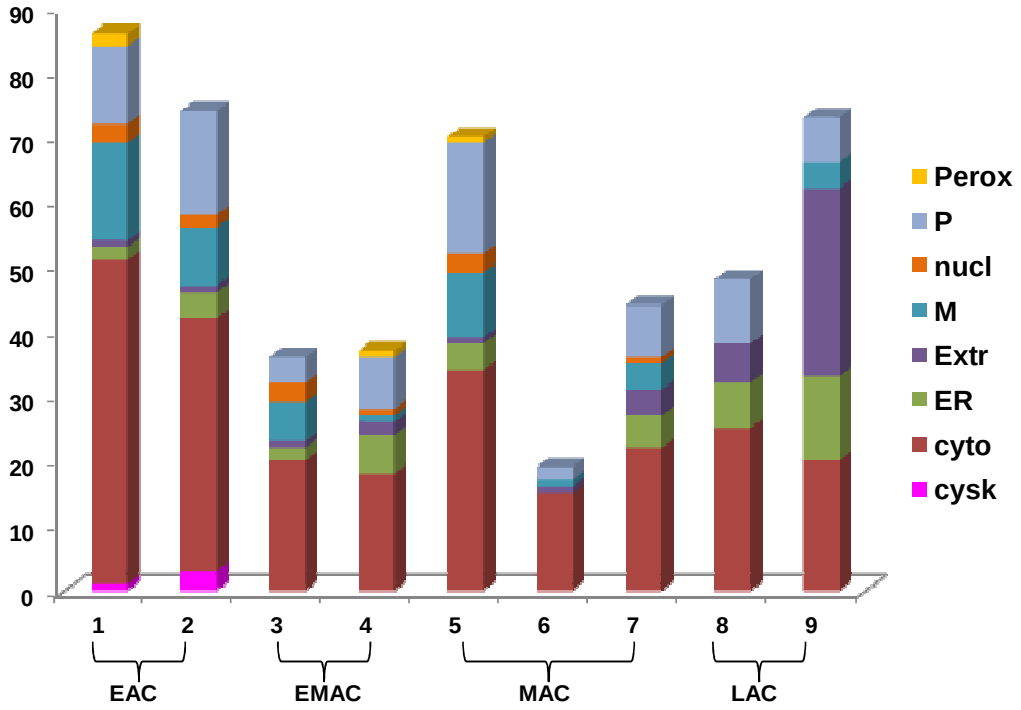




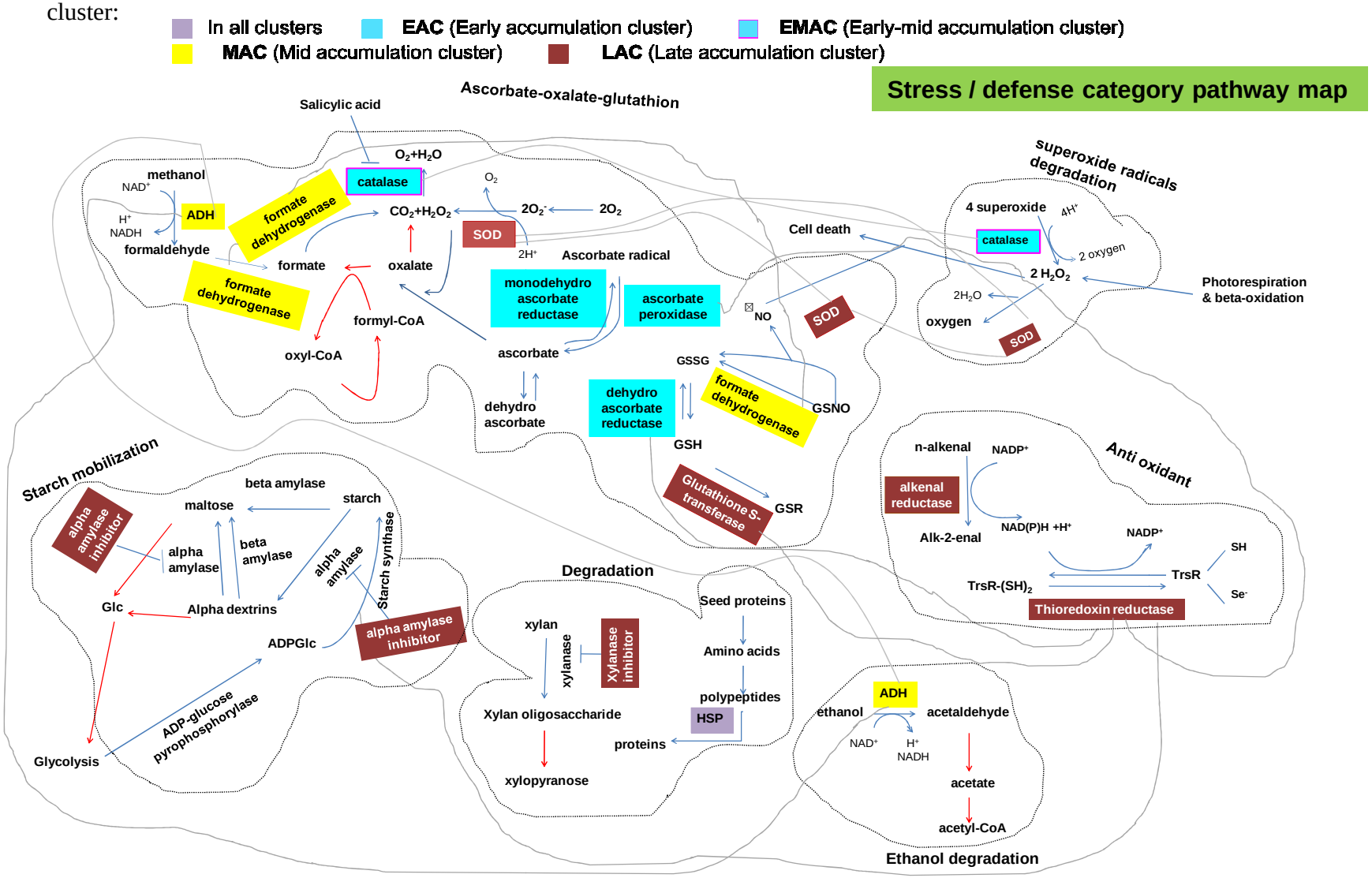
Supplemental Figure S3 : Hierarchical Clustering Analysis (HCA) of 950 significant spots.



Supplemental Figure S4: Predicted protein distribution in different sub-cellular localization of all identified proteins. X-axis represent profiles (P1 to P9), while y-axis represent number of proteins. Sub- cellular localizations are **cysk,** cytoskeleton; **cyto,** ctosol; **ER,** endoplasmic reticulum; **Extr,** extracellular region; **M,** mitochondria; **nucl,** nuclear; **P,** plastid; **Peroxi,** peroxisomal

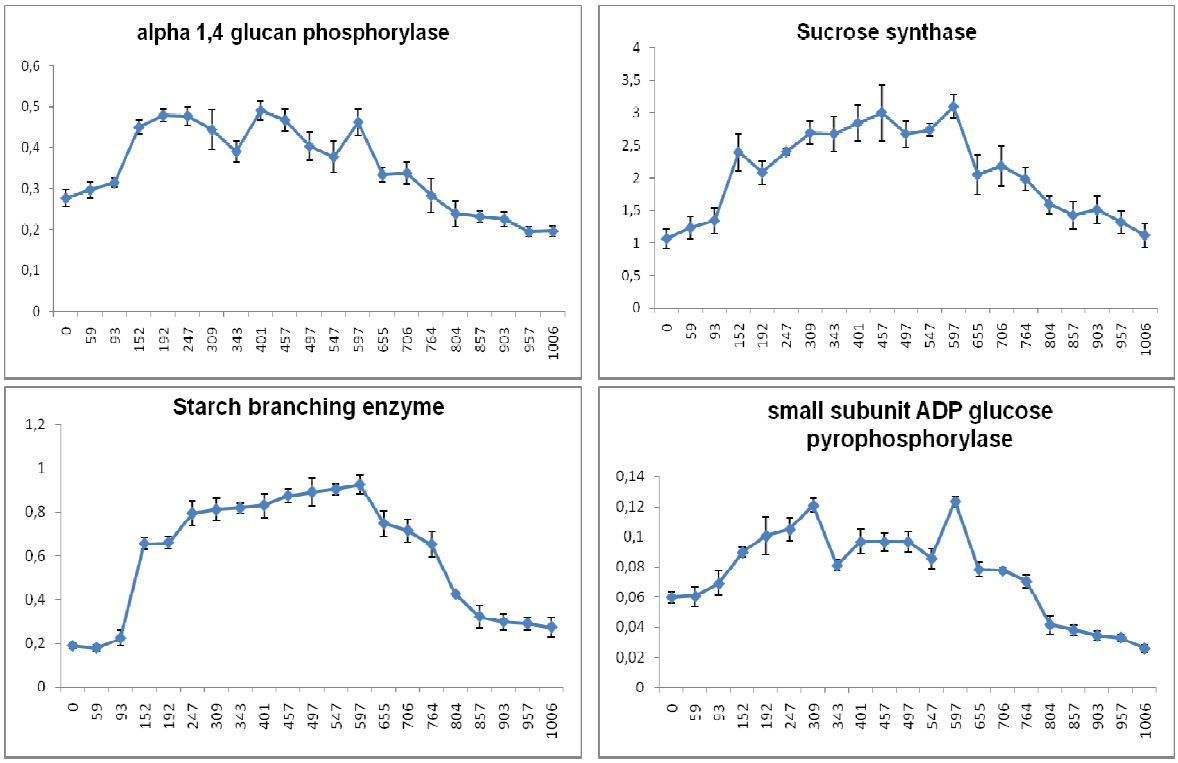


Supplemental Figure S5: Metabolic pathway of alg enzymes involved in Stress/ defense presented in developmental scale. Identified biochemical functions are shown by blue arrows while non identified are mentioned by red arrows. Colors represent the accumulation cluster:

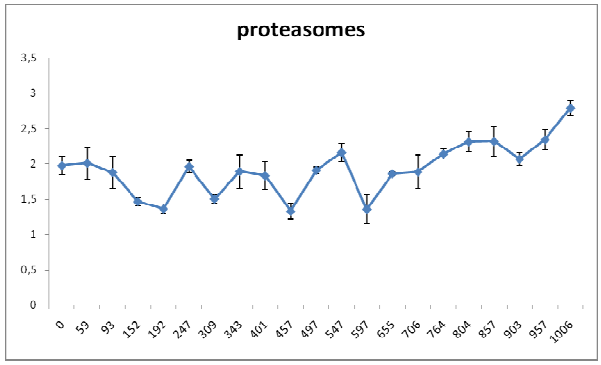


Supplemental Figure S6: a) Enzymes involved for starch biosynthesis and accumulation. **b)** Proteasome (PCD related protein) representative curve. Normalized volume is on vertical axis and stages of development (in °Cd) are on horizontal axis.

a



b



Chapter 4 (Article 3): An attempt to integrate proteome and transcriptome data: application to carbohydrate metabolism of wheat grain development

(data not yet published)

An attempt to integrate proteome and transcriptome data: application to carbohydrate metabolism of wheat grain development

Key words: *proteome, transcriptome, data integration, grain development, wheat*

Abbreviations: PR, proteins; TR, transcript; °Cd, degree days after pollination; CM, carbohydrate metabolism

Abstract

A preliminary approach for comparative study of proteome and transcriptome changes during grain development of *Triticum aestivum* (cultivar Récital) is described. All matched peptide sequences (varied from 3 to 30) of 145 identified proteins by LC-MS/MS were used for tblastn, to search corresponding transcripts. For 16 proteins no hits were found after tblastn search. Expression profiles were compared for proteome and transcriptome which resulted in 32% agreement between these two series of data. Divergent patterns were also seen, which is indicative of post-transcriptional or post-translational changes. We used the proteins identified in starchy endosperm for metabolism of carbohydrate. Enzymes involved in starch and sucrose biosynthesis showed the highest correlations between proteome-transcriptome profiles, among all other categories of carbohydrate metabolism. This study not only helped functional annotation of transcripts but also revealed the correlation between two “omics” data for the first time in wheat.

INTRODUCTION

Grain development in cereals is a complex process which involves a coordinated network of several metabolic pathways. Cereals are the main sources of staple starchy foods for humans, and wheat is one of the most widely grown, so knowledge of grain development in this species is important. Development of both the whole grain and of individual tissues has been studied using techniques including transcriptomic and proteomic to understand the underlying mechanisms (Wan et al. 2008, Nadaud et al. 2010, Robert et al. 2011). Earlier studies revealed that carbohydrate metabolism (CM) is of great importance for plant adaptation to varying environmental conditions (Guglielminetti et al. 1999).

Enzymes of carbohydrate metabolism (CM) are involved in important plant functions such as sucrose and starch synthesis, primary cell wall metabolism (pentose phosphate pathway), glycolysis and citrate cycle. During grain development, sucrose is converted to starch, the major storage carbohydrate in wheat, which accounts for 80% of mature grain weight, which is thus the main factor determining yield and important for both food and non-food industries (Tyson and ap Rees 1988, Poitrat 1999, James et al. 2003). As part of CM, starch synthesis is of prime importance since grain filling is mainly associated with this phenomenon (Yang et al. 2004), but other processes in which CM enzymes are involved, are also important in several ways. Glycolysis leads to carbohydrate degradation, the pentose phosphate pathway serves as a source of ribose 5-P and erythrose 4-P for synthesis of nucleotide and nucleic acid and for aromatic compounds respectively (Kelly and Latzko 1980). Study of CM during wheat grain development is important, as accumulation or depletion of different carbohydrates produce significant characteristics which can considerably change wheat properties (Kumar and Singh 1981).

Separate transcriptome and proteome analyses have been performed on developing seeds of wheat to understand the mechanisms and related pathways of CM (Wang et al. 2011, Tetlow et al. 1994, Dupont 2008). To follow precisely expression behavior during grain development, new strategies have been developed, such as pattern comparison of data issued from different “omics” techniques, more commonly transcriptomics, proteomics and metabolomics. Comparison of omics data, instead of analytical challenges, is helpful to provide more detailed insight into fundamental mechanisms involved during complex processes like development. Recently, some of these integrated “omics” studies have been reported for human, microorganisms and animals

(Wagner et al. 2006, Gómez-Pastor et al. 2010, Zhao et al. 2011). This strategy is also applied in plant species such as arabidopsis, medicago and maize (Gallardo et al. 2007, Hajduch et al. 2010, Casati et al. 2011). Proteome and transcriptome data comparison helped us to distinguish the processes regulated at transcriptome level and those controlled at the proteome level.

In the present study we monitored changes in enzymes of carbohydrate metabolism at protein and transcript level, with the aim of determining the relation among post-transcriptionally regulated transcript abundance and protein accumulation. Wheat (*Triticum aestivum*) grains at eleven stages, spanning the important developmental period related to CM enzymes (grain filling period) were analyzed using two major seed tissues (peripheral layers and endosperm).

MATERIAL AND METHODS

Plant Material

Wheat (*Triticum aestivum*) plants (cv Récital) were used for study, both of proteome and transcriptome analysis. The grains were cultivated and collected as described earlier (Tasleem-Tahir et al. 2012). We use the same eleven stages of grain development as were used to perform transcriptomic analysis: 152°Cd-706°Cd (degree days after pollination).

Transcriptomic Analysis

Transcriptomic analysis of endosperm grain development was performed earlier at eleven stages of development during grain filling (Romeuf 2010). Microarray profiling was performed by using 12x135k NimbleGen® cDNA microarray (Rustenholz et al. 2011), with two biological and two technical replicates. Nearly 35% of transcripts were expressed differentially during development, from 152°Cd-706°Cd.

Proteomic Analysis

For proteomic analysis, the steps involved from grain dissection to two-dimensional electrophoresis (2-DE) and identification using LC-MS/MS are described by Tasleem Tahir, A., 2012. The present comparison was made on 145 proteins involved in carbohydrate metabolism.

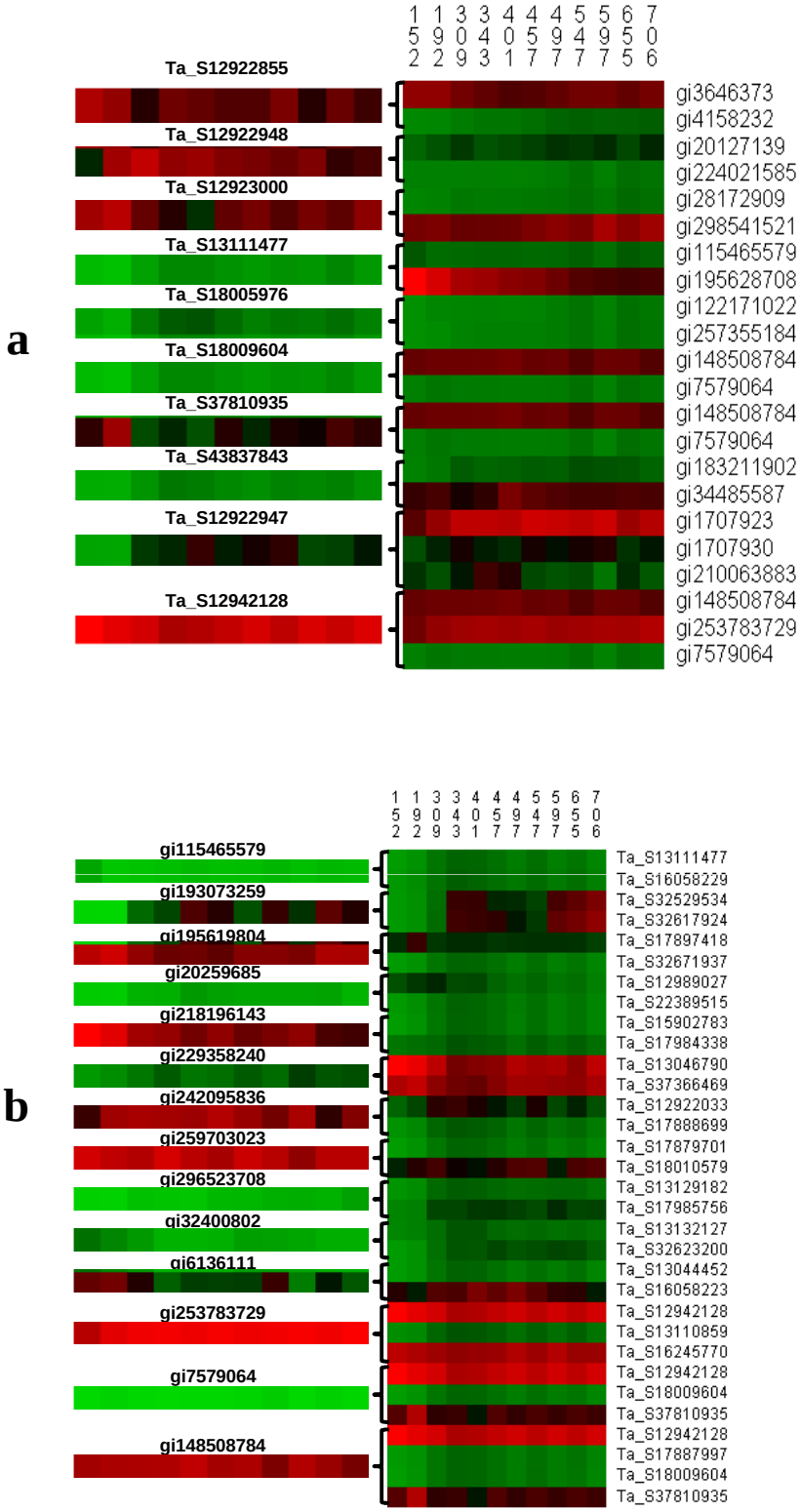


Figure 1: Expression profiles of proteins and transcripts. Permutmatrix was used to draw colour presentation, abundance decreased while moving from red to green color. Stages of development are shown on top in horizontal axis. **a)** 1 Transcript (on left) found in several proteins (on right), **b)** 1 protein (on left) expressed by several transcripts (on right).

Bioinformatics Analysis

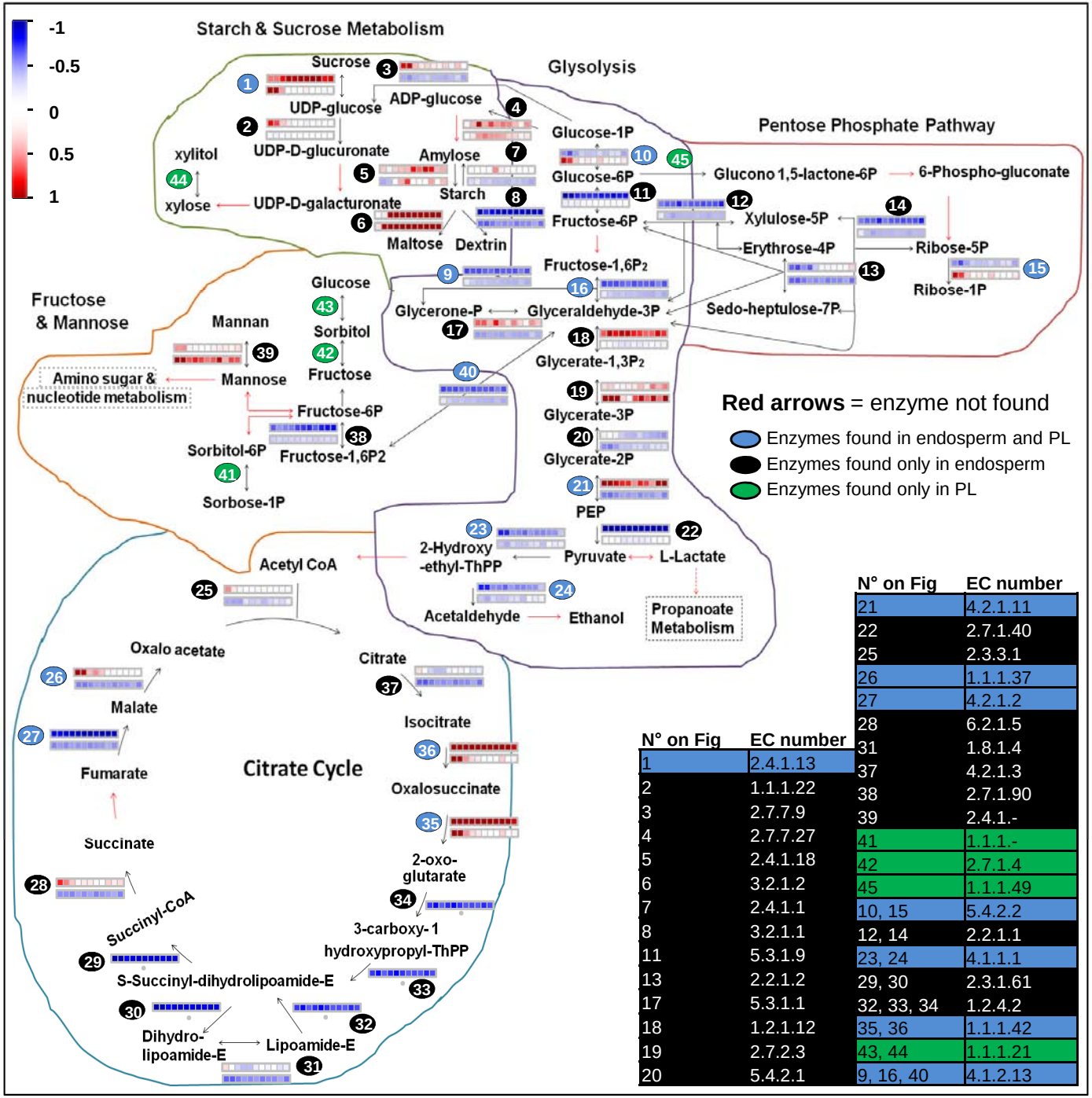
Protein identifications were considered valid when at least three peptides passed the significance criteria in LC-MS analysis and data base interrogation (Tasleem-Tahir et al. 2012). For CM category, the number of matched peptides for a single spot varied from 3 to 30. Only five proteins were validated with 3 peptides, and 58 were identified with more than 10 matched peptides.

We use individual sequences of all matched peptides, for each protein spot to search for sequence similarity against NCBI Wheat Unigene built 55 (http://www.ncbi.nlm.nih.gov/unigene), using tblastn program of BioEdit sequence alignment editor v7.1.3 (Hall 1999). The first hit of tblastn (with 100% identity and coverage) was considered as transcript (TR) sequence coded for given peptide. For each spot, TR that matched with a maximum number of peptides was validated (SuppInfo: Table 1 in Annex 2). In few cases, the same number of peptides matched to two or more TR, which resulted in more than one TR for a single protein spot. For these TR sequences, we used ClustalW2 alignment to determine their sequence similarity and the results were parsed using homemade Perl script. Further, these TR were translated to peptide sequences using homemade Perl script. For translation, we selected the same reading frame as for reference peptides. Finally, these translated sequences were compared with whole protein NCBI sequences to determine which TR was closest to our protein (SuppInfo: Table 2 in Annex 2). We also compared the translated sequences with our protein sequence, in the case where one TR was validated for one protein spot.

Proteome /Transcriptome Expression Pattern Comparison

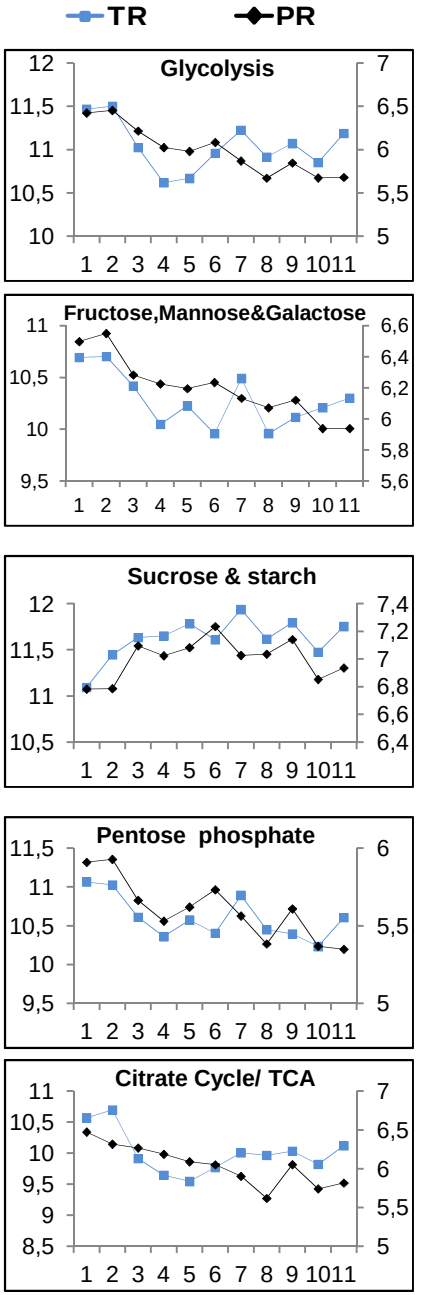
To compute PR and TR profile comparisons, $\log_2$ values for both data were used. Pearson correlation analysis was performed both for stage and spot variation. We used PermutMatrix (http://www.lirmm.fr/~caraux/PermutMatrix/) for pattern comparison of PR x TR pairs, which do not give one to one ratios (Figure 1). Mean center column values were used for expression presentation.

We assigned manually enzyme (EC) numbers to CM proteins identified in endosperm. To overview the CM enzymes, also expressed in grain peripheral layers, we used a previous study of PL during wheat grain development (Tasleem-Tahir et al. 2011). In this study, a total of 207 PL



N° on Fig	EC number
1	2.4.1.13
2	1.1.1.22
3	2.7.7.9
4	2.7.7.27
5	2.4.1.18
6	3.2.1.2
7	2.4.1.1
8	3.2.1.1
11	5.3.1.9
13	2.2.1.2
17	5.3.1.1
18	1.2.1.12
19	2.7.2.3
20	5.4.2.1

N° on Fig	EC number
21	4.2.1.11
22	2.7.1.40
25	2.3.3.1
26	1.1.1.37
27	4.2.1.2
28	6.2.1.5
31	1.8.1.4
37	4.2.1.3
38	2.7.1.90
39	2.4.1.-
41	1.1.1.-
42	2.7.1.4
45	1.1.1.49
10, 15	5.4.2.2
12, 14	2.2.1.1
23, 24	4.1.1.1
29, 30	2.3.1.61
32, 33, 34	1.2.4.2
35, 36	1.1.1.42
43, 44	1.1.1.21
9, 16, 40	4.1.2.13



proteins were identified, 27 of which were found to be involved in CM. After assignment of EC numbers to proteins of both grain tissues (SuppInfo Table 1 and 3 in Annex 2), we mapped them in five different CM categories with the help of Mapman (Usadel et al. 2005). These categories were: Glycolysis, Citrate cycle/TCA, Sucrose and starch synthesis, Pentose phosphate pathway and Fructose and mannose metabolism. Expression profiles were displayed for PR and TR of wheat grain endosperm data. For each enzyme, expression values of all corresponding proteins and transcripts were averaged to obtain a global expression for reference EC (Figure 2).

Prediction of Sub-cellular Localization

Sub-cellular localizations were predicted for endosperm proteins identified. We also predicted sub-cellular localizations for proteins of CM in peripheral layers using the same five programs as for endosperm: WoLF PSORT, Predotar, TargetP, YLoc and WegoLoc (Emanuelsson et al. 2000, Small et al. 2004, Horton et al. 2007, Briesemeister et al. 2010, Chi and Nam 2012). Sub-cellular localizations were validated, if at least three programs predicted the same localization of a protein at sub-cellular level.

RESULTS

Enzymes related to CM were monitored at eleven stages of grain development, at proteome and transcriptome level using 2-DE and NimbleGene microarray respectively. These stages covered the key period of grain filling, which is dominated by the main grain yield determinant process of starch synthesis.

Legend Figure 2: Enzymes of carbohyrate metabolism expressed in endosperm and grain peripheral layers. Each number corresponds to an enzyme as shown in side table presentation. Names of enzymes are given in SuppInfo Table 1 and Table 3. Colored bars represent expression of protein (upper bar) and transcript (lower bar), corresponding to reference enzyme. For each of five categories of CM, overall expression curves are shown on right of Mapman presentation. Trasncripts (TR) data is on primary axis with proteins (PR) on secondary axis.

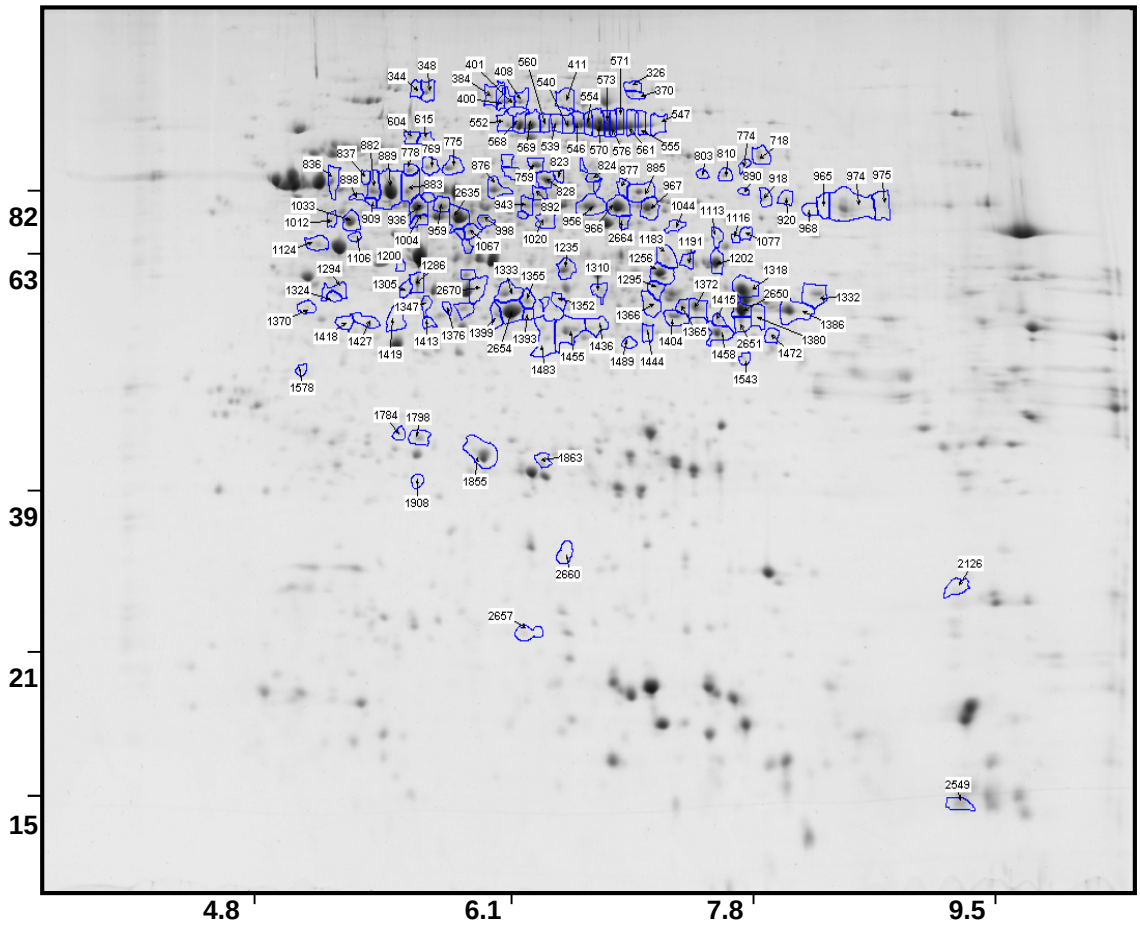


Figure 3: 2DE map of endosperm proteins during grain development. The outlined proteins are of carbohydrate metabolism at 457°Cd

sign	p Value	all pairs	Glycolysis	TCA*	Suc&Starch**	Pentose phosphate	Fruc,Mann, Galac***
Negative	<0.05	9	3		4	1	2
Negative	all	50	20	12	16	7	10
Positive	all	49	13	8	12	8	6
Positive	<0.05	37	10	3	17	5	5
Total		**145**	**46**	**23**	**49**	**21**	**23**
Significant		**46**	**13**	**3**	**21**	**6**	**7**
% Of significant pairs		**32**	**28**	**13**	**43**	**29**	**30**

Table 1: Pearson's correlation values, among proteins and transcripts expressed during grain development. *TCA=tricarboxylic acid cycle **=sucrose and starch synthesis, ***=Fructose, Mannose and Galactose metabolism.

A 2-DE map of CM included previously identified 145 proteins (PR) in endosperm and 27 in peripheral layers (Tasleem-Tahir et al. 2011, Tasleem-Tahir et al. 2012) (SuppInfo Table1 and 3 in Annex 2). Peripheral layer proteins were used only to map CM related enzymes in Figure 2, whereas, for all PR x TR comparisons, we only used proteins identified in endosperm. Their distribution on 2DE gel is shown in Figure 3. They were abundantly located in upper region of the gel, indicating that they are mostly of high molecular weight (>50 kDa) with only few in low molecular weight region. In addition, it was observed that during seed filling phase, proteins involved in CM were present in 5-8 pH range, on the gel.

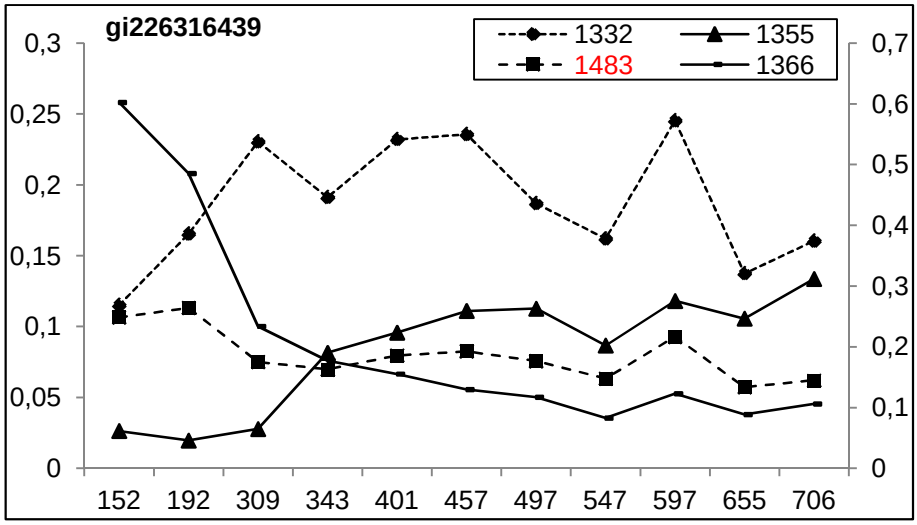
Each protein was represented by one to multiple spots (2-9) which resulted in 70 unique proteins. In our data, 32 proteins were identified in more than one spot and five were found in spots with opposing or significantly different expressions. Up to three different expression profiles for the same protein id were found in different spots: abundance at start of grain filling, with a plateau in the middle or towards end of grain filling or a spot with a peak at 706°Cd (Figure 4).

Tblastn Results

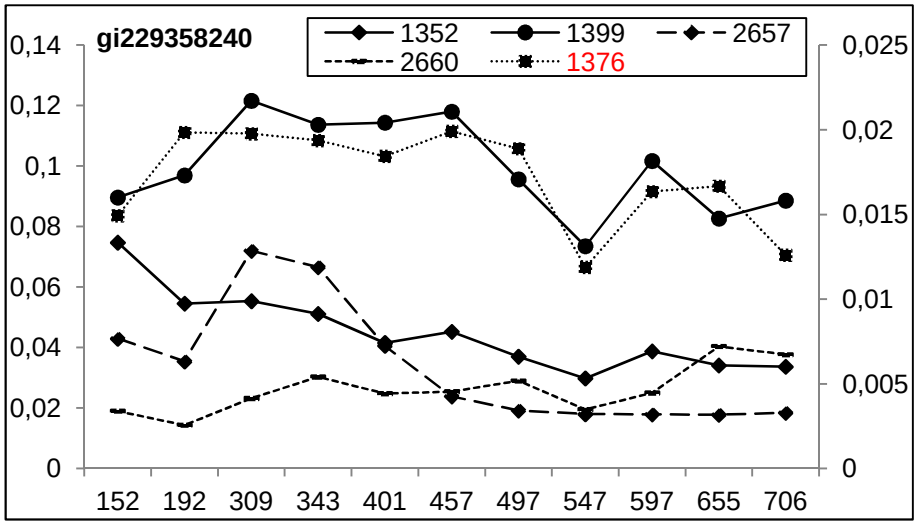
We found 135 transcripts (TR) in NCBI wheat Unigene, by performing tblastn. No single hit in Unigene transcript data was obtained for 16 spots of endosperm. For transcriptome and proteome data, one to one ratio or spots represented by multiple TR or vice versa were observed. 17 spots were represented by multiple TR (15 spots by 2 TR and 2 spots by 3TR, Figure 1).

For 1PR-1TR, scoring for sequence alignment of identified protein and translated sequence of corresponding TR resulted in highly significant scores varying between 80-100 except for one pair, (gi|401138 x Ta#S12923095) where a very low score (20) was observed. For all other cases (1PR-Many TR) we performed the same sequence alignment after translation of TR.

a



b



c

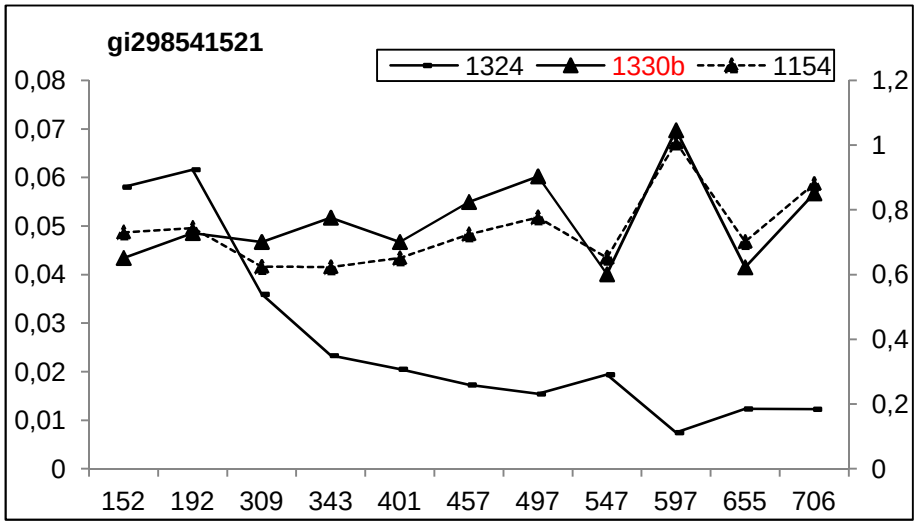


Figure 4- Presentation of some protein isoforms found in our analysis. Only spots with distinct expression are shown in the curves. **a:** fructose-bisphosphate aldolase **b:**cytosolic malate dehydrogenase **c:** unnamed protein product, after BlastP strong similarity with phosphoglycerate kinase, cytosolic was found. Stages of development are on horizontal axis and normalized volume is presented on vertical axis. Spots in red are presented on secondary axis.

Correlation Analysis of PR x TR data

To study the correlation between transcriptome and proteome data, a dynamic range of expression values for all identified proteins and their corresponding transcripts were investigated. The dynamic range of transcript signal intensities in the complete transcriptome was of higher magnitude, which would reflect a greater sensitivity of the microarray technology in the low expression range. Pearson's correlation coefficient statistics of proteins and transcript at individual stages of seed filling revealed that R^2 values increased with grain maturity. By employing correlation coefficient statistics to all proteins and their corresponding expression profiles of transcripts, nearly 32% of pairs showed significant behavior where only 6% of total pairs were negatively correlated and 26% with positive significant correlation (Table 1). Pearson's correlation coefficients were also computed for sub-categories of carbohydrate metabolism. Significant correlations among these categories were as follow: sucrose & starch: 43% (Neg=8%, Pos=35%), fructose, mannose & galactose: 30% (Neg=9%, Pos=21%), pentose phosphate metabolism: 29% (Neg=5%, Pos=24%), glycolysis: 28% (Neg=6%, Pos=22%) while TCA showed only a positive significant relation (13%) (Table 1). Some positive and negative significant correlations between PR and TR are shown on Figure 5. For negative correlations, two different cases were observed, PR increase and TR decrease or vice versa. These opposite behaviors intersect at different time points depending on the profiles which were correlated. Positive correlations were obtained for three different profiles, increase of both PR and TR with development, slight or significant decrease with end of grain filling.

In Figure 2 , mapping of CM proteins expressed in endosperm and PL, helped us to distinguish the processes specific to grain tissue. Enzymes involved in five main categories of CM were presented: glycolysis, citrate cycle/TCA, starch & sucrose metabolism, pentose phosphate pathway and fructose & mannose metabolism. Four enzymes were only found in PL, seven were observed in both tissues while 23 were found functional in endosperm. Details of enzymes shown in figure 2 are given in SuppInfo. Table 1 and 3 (Annex 2). Expression profiles of PR (upper row) and TR (lower row) are given only for CM proteins of endosperm. Each bar represents the stage of development from 152°Cd to 706°Cd. For a few spots, only expression profiles of PR are given, as we did not find the corresponding TR after tblastn.

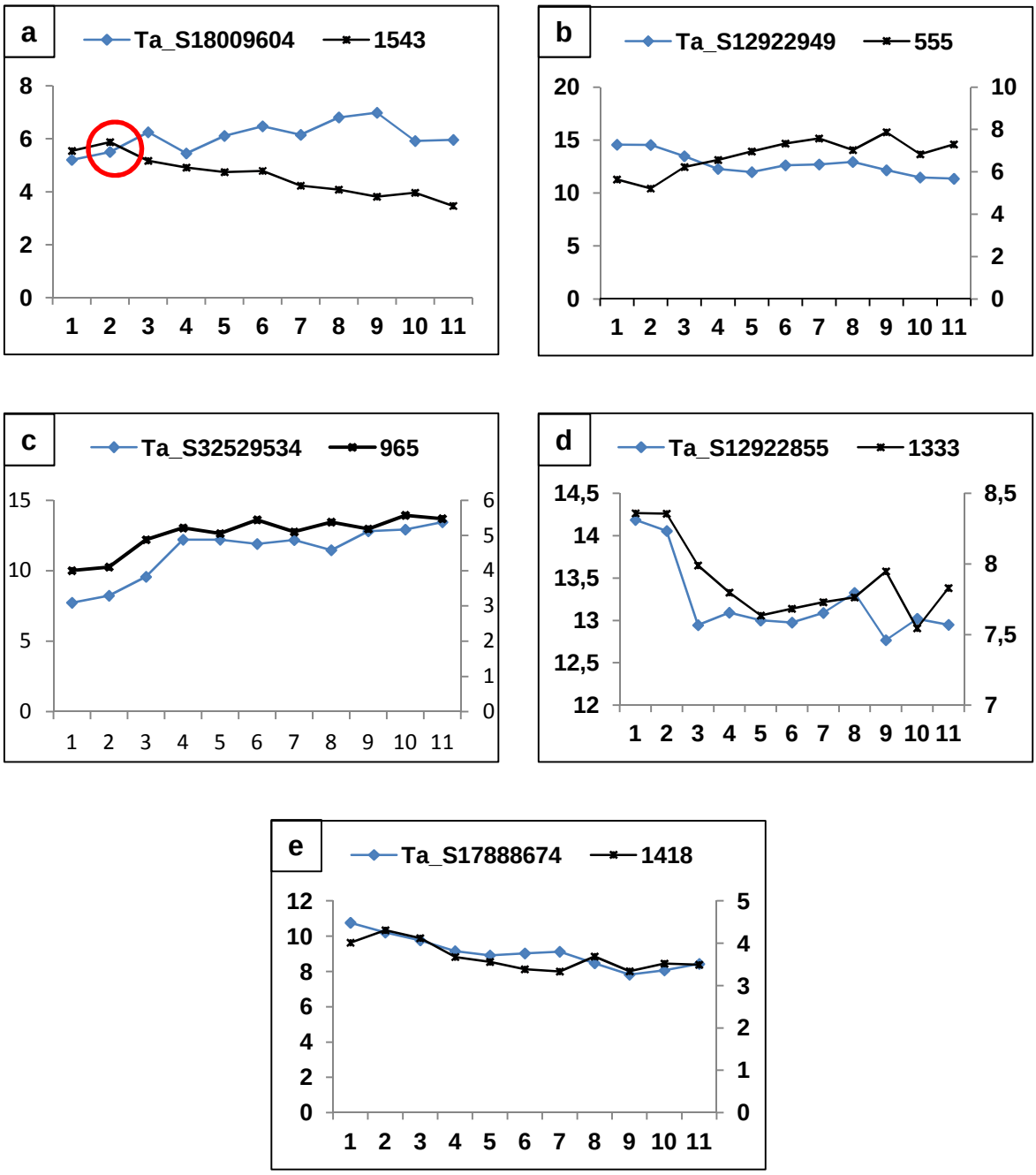


Figure 5: Significantly correlated pairs of protein (secondary axis) x transcript (primary axis). **a** and **b** are negatively correlated, **c**, **d** and **e** represent positively correlated pairs. Expression values are in log2 ratios and are on y-axis, while stages of development are on x-axis in the following sequence 152, 192, 309, 343, 401, 457, 497, 547, 597, 655 and 706°Cd.

Predicted Sub-cellular localization of CM proteins

Sub-cellular localization of identified CM proteins, varied during development. In a previous study using HCA, nine different protein expression profiles were observed according to their abundance at different stages of development from early grain growth period (P1) to grain maturity 1006°Cd (P9) (Tasleem-Tahir et al. 2012). In the present study, the grain filling period (152-706°Cd) is represented by proteins having expression profiles similar to P2 - P7. P1, P8 and P9 do not fall in the grain filling category as they represent protein abundance either at very early stages of development 0-93°C (P1) or at later stages of development (P8 and P9).

During grain filling period (P2-P7), proteins were found in five different locations, 46% of proteins were predicted in cytosol (Cyto), 30% in plastid (P), and 21% in mitochondria (M) (Figure 6). They were present in very small quantities in endoplasmic reticulum (ER, 2%), and extracellular regions (extr, 1%). In addition to grain filling period, proteins of CM were also found in P1, P8 and P9. They were scattered in 7 different locations in P1 (nucleus (nucl) peroxisome (perox), Cytoskeleton (cysk) in addition to five locations of grain filling period). In P8 and P9 they were found in the same localities as of grain filling period, but their abundance in ER and P was higher than that of P2 to P7.

Sub cellular localizations of PL proteins involved in CM, were also predicted using the same programs as used for endosperm proteins. They were found in four different localities, where their distribution was as follows: cyto 48%, plasma membrane (PM) 33%, M 11% an P 7%.

DISCUSSION

The present multidisciplinary study documents the dynamics of the transcriptomics and proteomics datasets, by unraveling the relationship among mRNA and protein patterns during seed filling period of wheat grain. Our focus was on metabolism of carbohydrate during grain development (152°Cd-706°Cd). Studies on *Arabidopsis* and *Medicago* showed that the genes involved in CM were among the most regulated in both species (Ruuska et al. 2002, Gallardo et al. 2007). During grain filling, uniform regulation of 27 proteins (out of 32) in multiple spots lead to the possibility that they may be generated from a single precursor protein through different

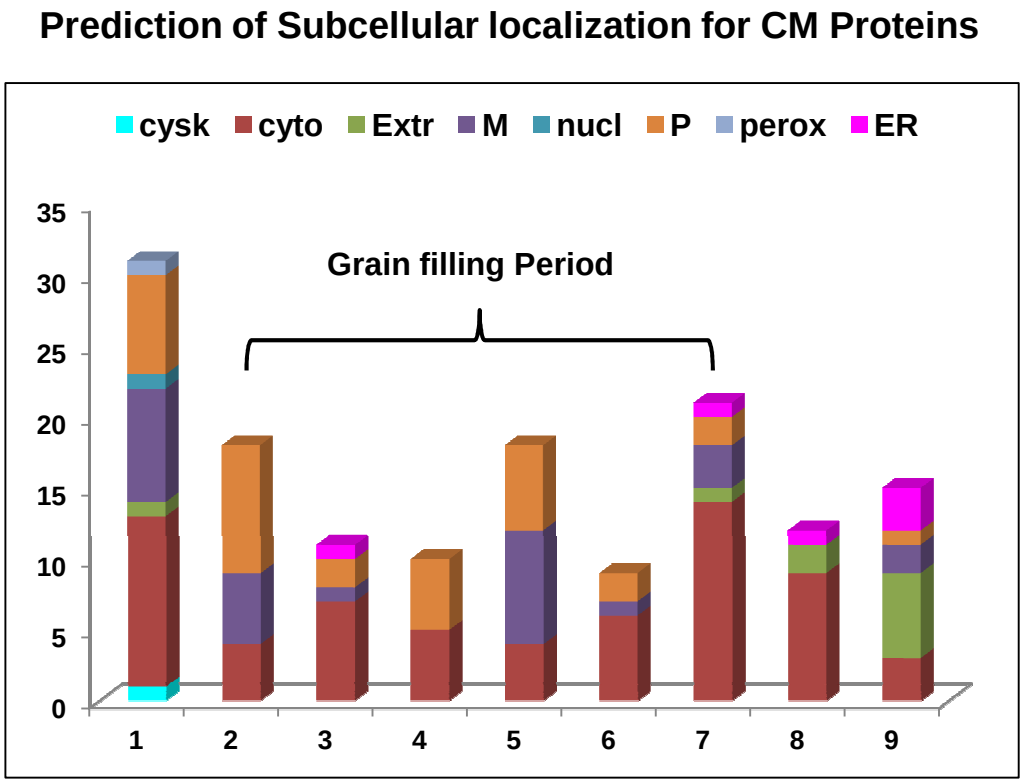


Figure 6: Predicted carbohydrate metabolism protein distribution in different subcellular localizations. *X*-axis represents profiles (P1–P9), while *y*-axis represents the number of identified proteins in corresponding profiles. **Cysk,** cytoskeleton; **Cyto**, cytosol; **ER**, endoplasmic reticulum; **Extr**, extra-cellular region; **M**, mitochondria; **nucl**, nuclear; **P**, plastid; **Perox**, peroxisome.

modifications such as alternative splicing of transcripts or posttranslational modifications. Numerous isoforms of CM proteins were also found in maize and rice grain development studies, which were from two to thirteen spots for the same proteins (Mechin et al. 2007, Xu et al. 2008). Moreover, a PRxTR comparative study on yeast suggested that presence of protein isoforms could result in disconcordance between TR and PR expression levels (Gómez-Pastor et al. 2010).

A slight decrease towards grain maturity was observed in overall expression of TR and PR. Wan et al reported that endosperm transcripts either have tendencies to increase to a plateau starting at 14 daa (~270°Cd) or some of them decrease with grain development (Wan et al 2008). These profiles are also evident in Figure 2. For sucrose and starch synthesis we observed a plateau, while for TCA and pentose phosphate enzymes, a significant decrease towards maturity was noticed. In addition to these two trends, we also observed that some TR first decrease and then increase slightly during development (glycolysis and fructose, mannose and galactose enzymes). Tendency to reach a plateau, observed for starch and sucrose enzymes, is considered as the characteristic curve for endosperm TR as starch accumulation is the dominant activity during grain filling. Noticeably, these TR plateau profiles were found similar to profiles of proteins involved in carbohydrate metabolic processes, which also showed a plateau (starch & sucrose synthesis) but in contrast with proteins involved in glycolysis; fructose, mannose & galactose; TCA; and pentose phosphate, for which decreasing profiles were observed. In Figure 2, enzymes were found in both grain tissues (endosperm and peripheral layers) among all categories of CM except fructose and mannose metabolism, where either enzyme was found only in endosperm or only in peripheral layers.

Remarkably, although mRNA intensities were more important than those of proteins, the variation in expression during development was higher in PR as compared to TR. This could result from higher sensitivity of proteome to various physiological conditions of the plant. In previous studies correlation among PR and TR vary from 20% to 70%, depending upon the system and approach used, but in general very low correlation values were found (Greenbaum et al. 2003, Watson et al. 2003, Schmidt et al. 2007, Pascal et al. 2008, Hajduch et al. 2010). Poor correlation could result from several factors: i) the analytical variability associated with global measurement technologies, such as differences in sensitivity, dynamic range, ambiguity in identification etc. ii) one gene does not usually code for a single protein as gene and protein

expressions are regulated separately by the cells. In a comparative PR x TR studies on human, it was found that 29% of genes from Affymetrix and 46% from Nimblegen do not match with sequence identifiers of protein data (Waters et al. 2006). This problem may arise due to inconsistencies in the gene and protein annotations used among different databases. iii) Although protein abundance depends on gene transcription rates, some additional control mechanisms, such as post-transcriptional mechanisms affecting mRNA stability, translational regulation and protein degradation also have an impact iv) a variety of post-translational modifications or proteolytic cleavage can alter the activity of proteins (Waters et al. 2006, Tan et al. 2009). In addition in wheat, the presence of three genomes increases the complexity of integration.

In our analysis, in addition to one-to-one relation among two datasets, we found one TR which corresponds to several spots or vice versa. After tblastn we chose the TR with 100% similarity to our peptide, this led to the hypothesis that these TRs could result from the same gene. Sequence comparison among translated TR sequences and those of proteins can give an indication of most probable TR, but as long as we do not have entire translated sequence information of corresponding gene, less reliability may result from the profile comparison. For our analysis when we encounter these cases (1TR-several spot or vice versa) all possibilities for correlation analysis were considered. Tblastn results of 1TR- several spots were obtained, most probably due to sequence similarity of peptides of these spots.

For the first time in wheat, an open-access data base named dbWFA (Web-Based Database for Functional Annotation) was developed (Vincent et al. 2012 data in press). Functional annotation of 35.5% of the transcripts for Wheat Gene Expression 12x135k NimbleGen array (Rustenholz et al. 2011) was performed. TR obtained through tblastn in our data set was of the same functional family in this database as found after annotation of TR on the basis of tblastn with identified proteins, in our experiment. In our data, information were not only precise (as annotation was made on experimental basis), but some additional TR (not annotated in this database) were also annotated.

This approach documented information concerning PR and TR and correlations between their expression profiles. We used proteins involved in carbohydrate metabolism as an example for application of this strategy. It was noticed that among five categories of carbohydrate metabolism studied, sucrose and starch synthesis showed maximum significant correlations. Moreover, the

enzymatic profile study during development could be useful; to determine whether the corresponding reaction is controlled at PR or TR level. Further, to obtain a comprehensive overview of grain development, it would be useful to study other metabolisms and their corresponding profiles by integration of proteomic and transcriptomic data.

REFERENCES

Briesemeister, S., J. Rahnenführer, and O. Kohlbacher. 2010. YLoc—an interpretable web server for predicting subcellular localization. Nucleic Acids Research 38: W497-W502.

Casati, P., M. Campi, D. J. Morrow, J. Fernandes, and V. Walbot. 2011. Transcriptomic, proteomic and metabolomic analysis of maize responses to UV-B: comparison of greenhouse and field growth conditions. Plant signaling & behavior 6: 1146-1153.

Chi, S.-M., and D. Nam. 2012. WegoLoc: accurate prediction of protein subcellular localization using weighted Gene Ontology terms. Bioinformatics 28: 1028-1030.

Dupont, F. 2008. Metabolic pathways of the wheat (*Triticum aestivum*) endosperm amyloplast revealed by proteomics. BMC plant biology 8: 39.

Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology 300: 1005-1016.

Gallardo, K., C. Firnhaber, H. Zuber, D. Héricher, M. Belghazi, C. Henry, H. Küster, and R. Thompson. 2007. A Combined Proteome and Transcriptome Analysis of Developing *Medicago truncatula* Seeds. Molecular & Cellular Proteomics 6: 2165-2179.

Gómez-Pastor, R., R. Pérez-Torrado, E. Cabiscol, and E. Matallana. 2010. Transcriptomic and proteomic insights of the wine yeast biomass propagation process. FEMS Yeast Research 10: 870-884.

Greenbaum, D., C. Colangelo, K. Williams, and M. Gerstein. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biology 4: 117.

Guglielminetti, L., E. Loreti, P. Perata, and A. Alpi. 1999. Sucrose Synthesis in Cereal Grains under Oxygen Deprivation. Journal of Plant Research 112: 353-359.

Hajduch, M., L. B. Hearne, J. A. Miernyk, J. E. Casteel, T. Joshi, G. K. Agrawal, Z. Song, M. Zhou, D. Xu, and J. J. Thelen. 2010. Systems analysis of seed filling in *Arabidopsis*: using general linear modeling to assess concordance of transcript and protein expression. Plant physiology 152: 2078-2087.

Hall, T. A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser ;41:95-8

Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res 35: W585-587.

James, M. G., K. Denyer, and A. M. Myers. 2003. Starch synthesis in the cereal endosperm. Current Opinion in Plant Biology 6: 215-222.

Kelly, G. J., Latzko, E. 1980. In "The Biochemistry of Plants"; Stumpf, P. K.; Conn, E. E., Eds; Academic Press: New York and London. Vol. 1, p p 183-208.

Kumar, R., and R. Singh. 1981. Free sugars and their relationship with grain size and starch content in developing wheat grains. Journal of the Science of Food and Agriculture 32: 229-234.

Mechin, V., C. Thevenot, M. Le Guilloux, J. L. Prioul, and C. Damerval. 2007. Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. Plant physiology 143: 1203-1219.

Nadaud, I., C. Girousse, C. Debiton, C. Chambon, M. F. Bouzidi, P. Martre, and G. Branlard. 2010. Proteomic and morphological analysis of early stages of wheat grain development. PROTEOMICS 10: 2901-2910.

Pascal, L., L. True, D. Campbell, E. Deutsch, M. Risk, I. Coleman, L. Eichner, P. Nelson, and A. Liu. 2008. Correlation of mRNA and protein levels: Cell type-specific gene expression of cluster designation antigens in the prostate. BMC genomics 9: 246.

Poitrat, E. 1999. The potential of liquid biofuels in France. Renewable Energy 16: 1084-1089.

Robert, P., F. Jamme, C. Barron, B. Bouchet, L. Saulnier, P. Dumas, and F. Guillon. 2011. Change in wall composition of transfer and aleurone cells during wheat grain development. Planta 233: 393-406.

Romeuf, I., 2010. Identification in silico des facteurs de transcription du blé tendre (*Triticum aestivum*) et mise en évidence des facteurs de transcription impliqués dans la synthèse des protéines de réserve. Ph.D. thesis, Université Blaise Pascal, Clermont-Ferrand, France.pp205.

Rustenholz, C., F. Choulet, C. Laugier, J. Šafář, H. Šimková, J. Doležel, F. Magni, S. Scalabrin, F. Cattonaro, S. Vautrin, A. Bellec, H. Bergès, C. Feuillet, and E. Paux. 2011. A 3000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. Plant physiology **157:** 1596-1608.

Ruuska, S. A., T. Girke, C. Benning, and J. B. Ohlrogge. 2002. Contrapuntal Networks of Gene Expression during Arabidopsis Seed Filling. The Plant Cell Online 14: 1191-1206.

Schmidt, M. W., A. Houseman, A. R. Ivanov, and D. A. Wolf. 2007. Comparative proteomic and transcriptomic profiling of the fission yeast Schizosaccharomyces pombe. Molecular systems biology 3: 79.

Small, I., N. Peeters, F. Legeai, and C. Lurin. 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. PROTEOMICS 4: 1581-1590.

Tan, C., A. Salim, A. Ploner, J. Lehtio, K. Chia, and Y. Pawitan. 2009. Correlating gene and protein expression data using Correlated Factor Analysis. BMC Bioinformatics 10: 272.

Tasleem-Tahir, A., I. Nadaud, C. Chambon, and G. Branlard. 2012. Expression Profiling of Starchy Endosperm Metabolic Proteins at 21 Stages of Wheat Grain Development. Journal of Proteome Research 11: 2754-2773.

Tasleem-Tahir, A., I. Nadaud, C. Girousse, P. Martre, D. Marion, and G. Branlard. 2011. Proteomic analysis of peripheral layers during wheat (*Triticum aestivum L.*) grain development. PROTEOMICS 11: 371-379.

Tetlow, I. J., K. J. Blissett, and M. J. Emes. 1994. Starch synthesis and carbohydrate oxidation in amyloplasts from developing wheat endosperm. Planta 194: 454-460.

Tyson, R. H., and T. ap Rees. 1988. Starch synthesis by isolated amyloplasts from wheat endosperm. Planta 175: 33-38.

Usadel, B., A. Nagel, O. Thimm, H. Redestig, O. E. Blaesing, N. Palacios-Rojas, J. Selbig, J. Hannemann, M. C. Piques, D. Steinhauser, W.-R. Scheible, Y. Gibon, R. Morcuende, D. Weicht, S. Meyer, and M. Stitt. 2005. Extension of the Visualization Tool MapMan

to Allow Statistical Analysis of Arrays, Display of Coresponding Genes, and Comparison with Known Responses. Plant physiology 138: 1195-1204.

Vincent, J., Dai, Z.W., Ravel, C., Agier, M., Mouzeyar, S., Bouzidi, M.F., and Martre P. dbWFA: a web-based database for functional annotation of wheat transcripts. In press.

Wagner, W., R. E. Feldmann Jr, A. Seckinger, M. H. Maurer, F. Wein, J. Blake, U. Krause, A. Kalenka, H. F. Bürgers, R. Saffrich, P. Wuchter, W. Kuschinsky, and A. D. Ho. 2006. The heterogeneity of human mesenchymal stem cell preparations—Evidence from simultaneous analysis of proteomes and transcriptomes. Experimental Hematology 34: 536-548.

Wan, Y., R. L. Poole, A. K. Huttly, C. Toscano-Underwood, K. Feeney, S. Welham, M. J. Gooding, C. Mills, K. J. Edwards, P. R. Shewry, and R. A. Mitchell. 2008. Transcriptome analysis of grain development in hexaploid wheat. BMC genomics 9: 121.

Wang, Z., W. Li, J. Qi, P. Shi, and Y. Yin. 2011. Starch accumulation, activities of key enzyme and gene expression in starch synthesis of wheat endosperm with different starch contents. Journal of Food Science and Technology: 1-11.

Waters, K. M., J. G. Pounds, and B. D. Thrall. 2006. Data merging for integrated microarray and proteomic analysis. Briefings in Functional Genomics & Proteomics 5: 261-272.

Watson, B. S., V. S. Asirvatham, L. Wang, and L. W. Sumner. 2003. Mapping the Proteome of Barrel Medic (Medicago truncatula). Plant physiology 131: 1104-1123.

Xu, S. B., T. Li, Z. Y. Deng, K. Chong, Y. Xue, and T. Wang. 2008. Dynamic Proteomic Analysis Reveals a Switch between Central Carbon Metabolism and Alcoholic Fermentation in Rice Filling Grains. Plant physiology 148: 908-925.

Yang, J., J. Zhang, Z. Wang, G. Xu, and Q. Zhu. 2004. Activities of Key Enzymes in Sucrose-to-Starch Conversion in Wheat Grains Subjected to Water Deficit during Grain Filling. Plant physiology 135: 1621-1629.

Zhao, Y., P. Xie, and H. Fan. 2011. Genomic Profiling of MicroRNAs and Proteomics Reveals an Early Molecular Alteration Associated with Tumorigenesis Induced by MC-LR in Mice. Environmental science & technology 46: 34-41.

Chapter 5: Comparative proteome study of Peripheral layers and Starchy Endosperm during grain development

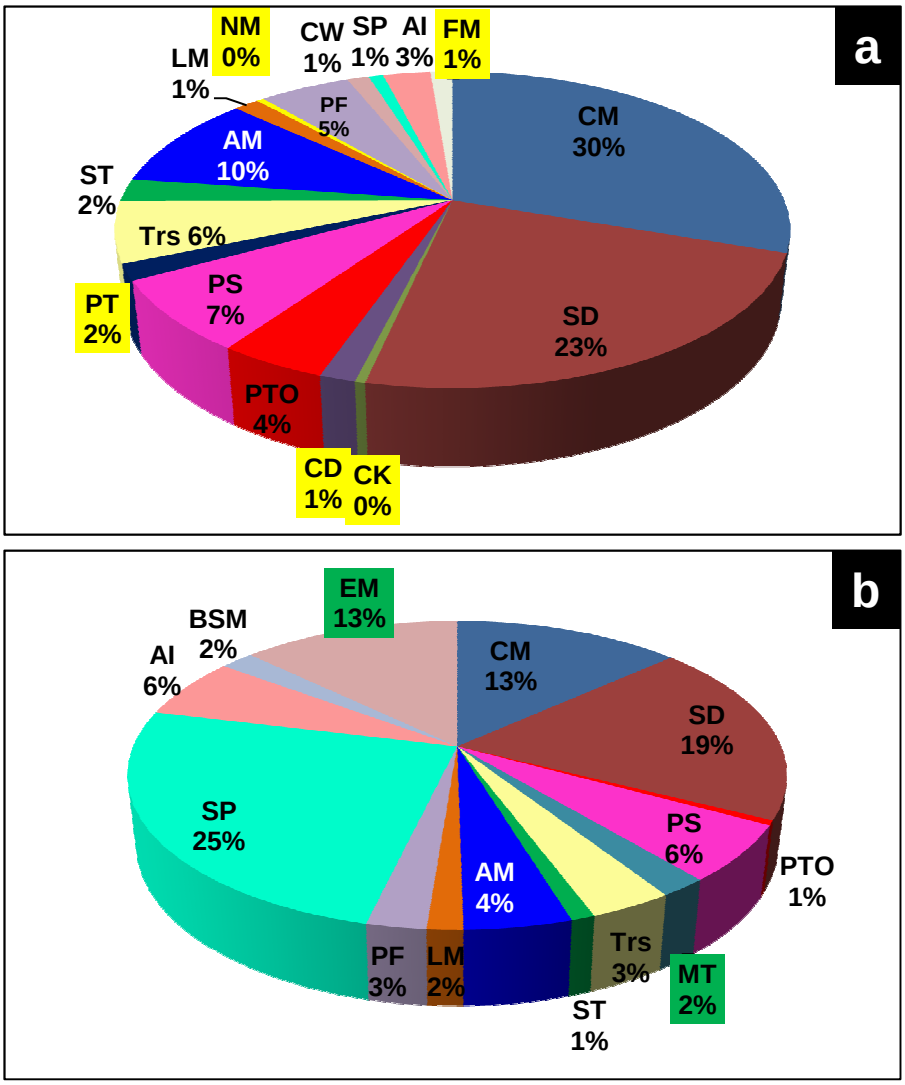


Figure 23: Metabolic proteins identified in grain tissues. **a)** Proteins identified in starchy endosperm (ESM) **b)** proteins identified in peripheral layers (PL).Yellow tagged functional categories (CK, CD, NM, PT and FM) were only identified in ESM, while those tagged in green (MT, EM) were identifed only in PL. Protein identifications as unknown function are not mentioned in these pie charts.

Rank on basis of abundance	ESM	PL
1	CM	SP
2	SD	SD
3	AM	CM, EM
4	PS	AI, PS
5	Trs	AM
6	PF	Trs, PF
7	PTO	BSM, MT, LM
8	AI	ST, PTO
9	ST, PT	
10	CD, LM, CW, FM, SP	
11	CK, NM	

Table 11: Functional distribution in starchy endosperm (ESM) and peripheral layers (PL) in descending order.

We studied grain development, from fertilization to maturity, with the objective of unraveling the mechanisms responsible for this complex and precisely planned process. Analyses were performed separately on dissected grain tissues: peripheral layers and endosperm. This proved to be advantageous, as it enabled: i) analysis of tissue-specific protein patterns, ii) reduced complexity in comparison to whole seed extracts, making it easier to visualise and identify less abundant proteins (Finnie and Svensson 2003). We studied metabolic proteins in both tissues, to make possible comparisons of their profiles. For PL, 109-700°Cd grains were dissected, while for ESM, we dissected the grains from 152°Cd to 1006°Cd.

Here, we will discuss similarities and differences among these tissues, based on the proteins identified and their expression profiles.

1 Metabolic Proteins identified in Grain Tissues

Total proteins identified in PL and ESM are presented in Figure 23. The dominant functional categories of these tissues were different. In Table 11 they are presented in decreasing order of abundance. Remarkably, storage proteins (25%) were the dominant category in PL but only showed minor presence in ESM (only 1%). Stress/defense (19%), carbohydrate and energy metabolism (13% each) followed the largest category, of storage proteins, in peripheral layers.

In starchy endosperm, the proteins of carbohydrate metabolism were found to be the most abundant category, representing nearly 30% of total identified proteins, followed by stress/defense (23%) and amino acid metabolism (10%).

Functional categories tagged in green (EM, MT) were only identified in PL. The presence of abundant EM related proteins, only in PL, was mainly due to photosynthetic activities of the green pericarp (part of PL), which is absent from endosperm. Four lipoproteins were identified and classed in MT category. They were only identified in PL, as these layers are rich in lipids as compared to ESM. CK and CD, yellow tagged categories and characteristic of early grain development stages, were only identified in ESM.

2 Expression Profiles of PL and ESM proteins

In PL we observed nearly 800 spots that varied significantly over fifteen stages of development, while, for ESM, 950 spots varied significantly over twenty one stages. Using HCA, five different expression profiles (P1-P5) were distinct in PL and nine profiles (P1-P9) in ESM (Chapter 2 and Chapter 3). Among these profiles, only profile 5 of PL and profile 9 of ESM showed similar variation patterns, i:e increase from stage six towards maturity. In profile 5 of PL, the dominant class was storage proteins, while in P9 of ESM, the stress /defense category was dominant. Remarkably going from mid to later stages of development, in peripheral layers, greatest abundance moved from energy metabolism (P4) to storage proteins (P5) while, in endosperm, the change was from carbohydrate metabolism (P7) to stress/ defense proteins (P9).

3 Functional Profiles of PL and ESM

We compared the evolutionary profiles of functional categories that were found common to both tissues. Grain development was monitored for each 50°Cd; in PL we analyzed fifteen stages while six additional stages were added for analysis of starchy endosperm tissue. In the publication concerning PL, the expression profiles were presented as "quantity of protein µg/PL of grain", while for ESM, they were in normalized volume values. To homogenize the scale, we present here normalized volume values for both tissues. The main and sub categories are given below.

3.1 METABOLISM

In this category, CM, AM and LM were found in both tissues, while proteins involved in FM and NM were identified only in ESM, and those related to biosynthesis of secondary metabolites only in PL (Figure 23).

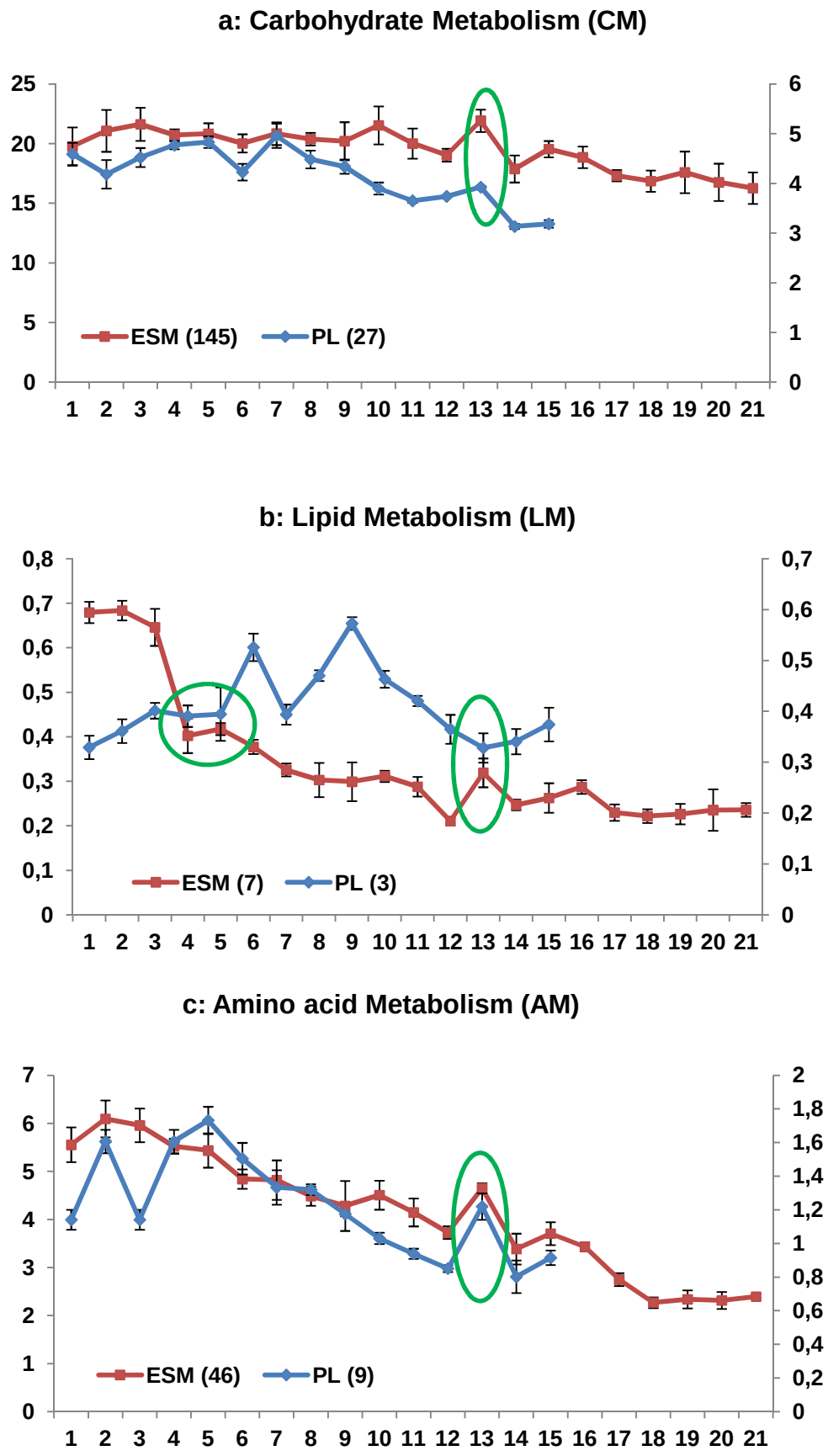


Figure 24: Protein expression profiles of categories, classed under metabolism (**a-c**). Stages of development are on x-axis, while normalized volume values are presented on y-axis, for starchy endoseprm (ESM) on primary while for pripheral layers (PL) on secondary axis. Number of identified proteins for each category are given in brackets.

a) Carbohydrate Metabolism (CM)

We identified 145 proteins in ESM and 27 in PL related to this category. These proteins tended to decrease towards grain maturity in both tissues (Figure 24 a). This decrease was linear and slight with an up-regulation at stage 13 in both tissues. The proteins of CM category are discussed also in Chapter 4.

b) Lipid Metabolism (LM)

There were 3 LM proteins identified in PL whereas 7 in ESM. Notably, these proteins were abundant at three early stages in ESM in contrast to very weak presence in PL. Their expressions intersected at stage 4, followed by a decrease in ESM and an increase until stage 9 in PL. Similar to CM proteins expression, change of behaviour at stage 13 was found as compared to adjacent stages. Interestingly, like CM, at stage 13 the behaviour of LM proteins in ESM also differed in comparison to adjacent stages (Figure 24 b). No proteins were found common between PL (aspartic proteinase and oleosin) and ESM (enoyl-ACP reductase, stearoyl-ACP desaturase and acyl carrier proteins).

c) Amino acid Metabolism (AM)

Nine proteins identified in PL and forty six of ESM showed parallel expression through all stages, except at stage 3 in PL, when protein abundance decreased unexpectedly (Figure 24 c). All proteins identified in peripheral layers were also found in endosperm, while some additional proteins also identified in endosperm.

3.2 GENETIC INFORMATION AND PROCESSING

a) Protein Synthesis (PS)

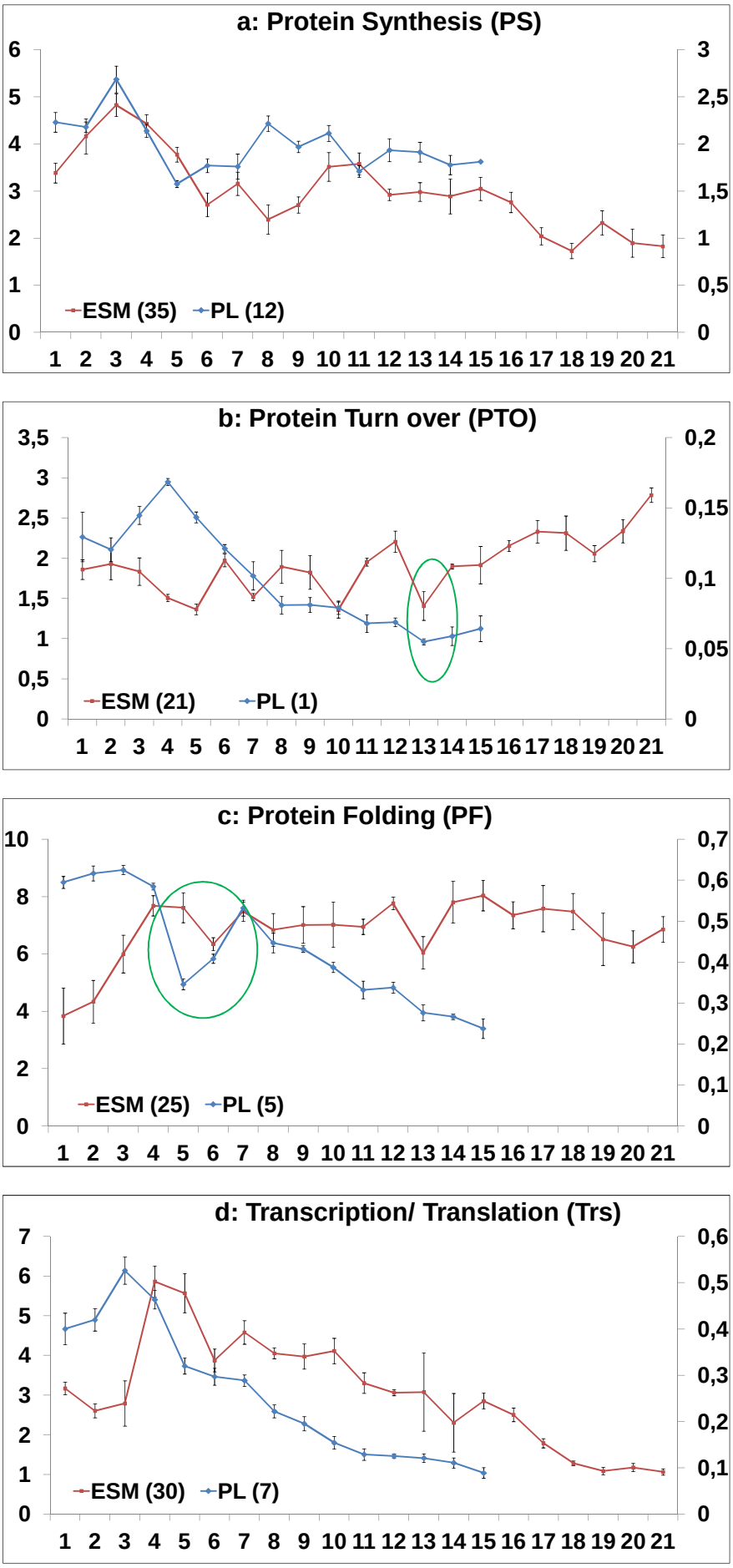


Figure 25: Protein expression profiles of categories classed under genetic information and processing (**a-d**). Stages of development are on x-axis, while normalized volume values are presented on y-axis, for starchy endoseprm (ESM) on primary axis while for pripheral layers (PL) on secondary axis. Number of identified proteins for each category are in brackets.

Among 12 PS proteins of PL and 35 of ESM, HSP 70 and secretary proteins were common in these tissues, while low molecular weight HSP (16-17 KDa) were found only in ESM. Expression profiles moved nearly parallel until stage 7, after this stage the behavior was opposite with an intersection at stage 11 (Figure 25 a). In PL, as our analysis was only till stage 15, we have no data after this time for these layers. It is possible that like ESM proteins, they continue to decrease after physiological maturity.

b) Protein Turn over (PTO)

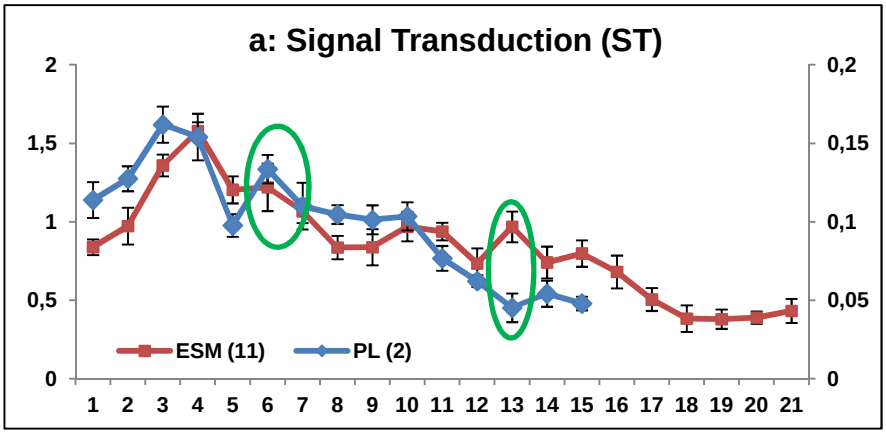
Only one protein of PTO identified in PL peaked at stage 4 and then decreased significantly. On the other hand, in ESM, 21 identified proteins of PTO were more or less constant until stage 13 and then their abundance increased with peak at stage 21 (Figure 25 b). Remarkably, PL proteins also showed a tendency of increased abundance after stage 13. In this class, proteasome was found to be common but with different isoforms, while ubiquitin was only identified in ESM.

c) Protein Folding (PF)

We identified 5 proteins of PF in outer layers but 25 in grain ESM. These proteins showed opposite expression behavior in PL and ESM, except at stages 6 and 7, where their profiles were parallel to each other (Figure 25 c). PDI and chaperonin were common in both grain parts, but with different isoforms, while cyclophilin were only identified in ESM. This protein folding function was found significantly active in PL and in ESM at different time points.

d) Transcription/ Translation (Trs)

These proteins in PL (7) and ESM (30) were abundant at early stages of development. They peaked earlier in PL (at stage 3) and later in ESM (at stage 4 and 5), and then continued to decrease with development (Figure 25 d). Translation elongation and initiation factors were the proteins identified in common, while ribo-nucleoproteins were only identified in PL.



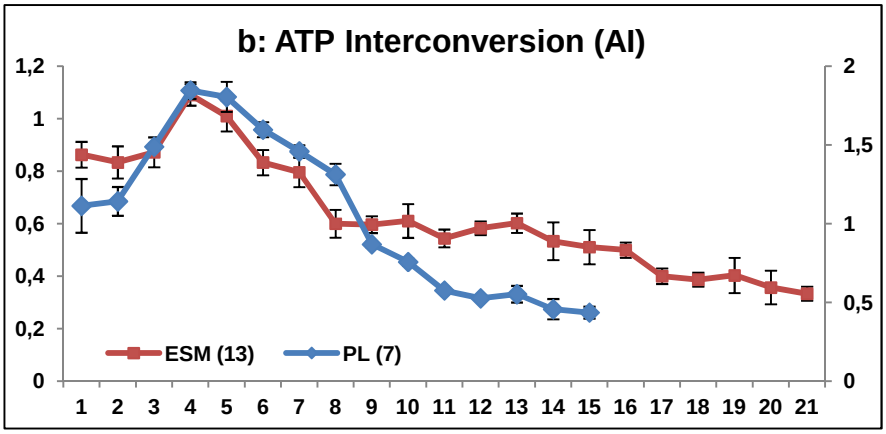


Figure 26: Protein expression profiles of categories classed under Environmental Information and Processing (**a-b**). Stages of development are on x-axis, while normalized volume values are presented on y-axis, for starchy endoseprm (ESM) on primary while for pripheral layers (PL) on secondary axis. Number of identified proteins for each category are given in brackets.

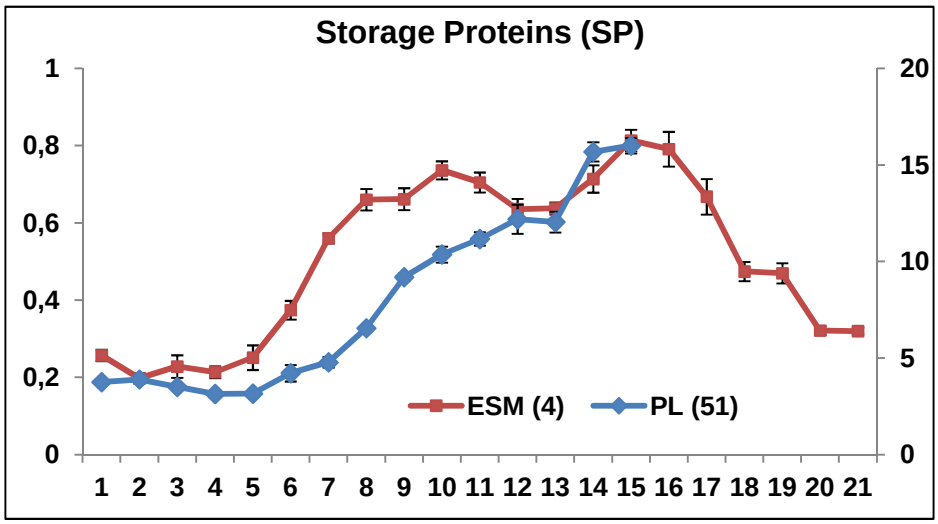


Figure 27: Protein expression profiles of storage proteins. Stages of development are on x-axis, while normalized volume values are presented on y-axis, for starchy endoseprm (ESM) on primary while for pripheral layers (PL) on secondary axis. Number of identified proteins for each category are given in brackets.

3.3 ENVIRONMENTAL INFORMATION AND PROCESSING

a) Signal Transduction (ST)

The peak of the 2 ST related proteins in PL was one stage earlier than that of ESM (11 proteins), the same delayed peak was observed for Trs related proteins (Figure 25 d). After a peaked expression in PL (at stage 3) and ESM (at stage 4), these proteins decreased with development, showing slightly different behaviors at stage 6 and 13 compared with their adjacent stages (Figure 26 a). Different isoforms of 14-3-3- like proteins were found in both tissues, while some other proteins such as annexin and GTP binding were identified only in ESM.

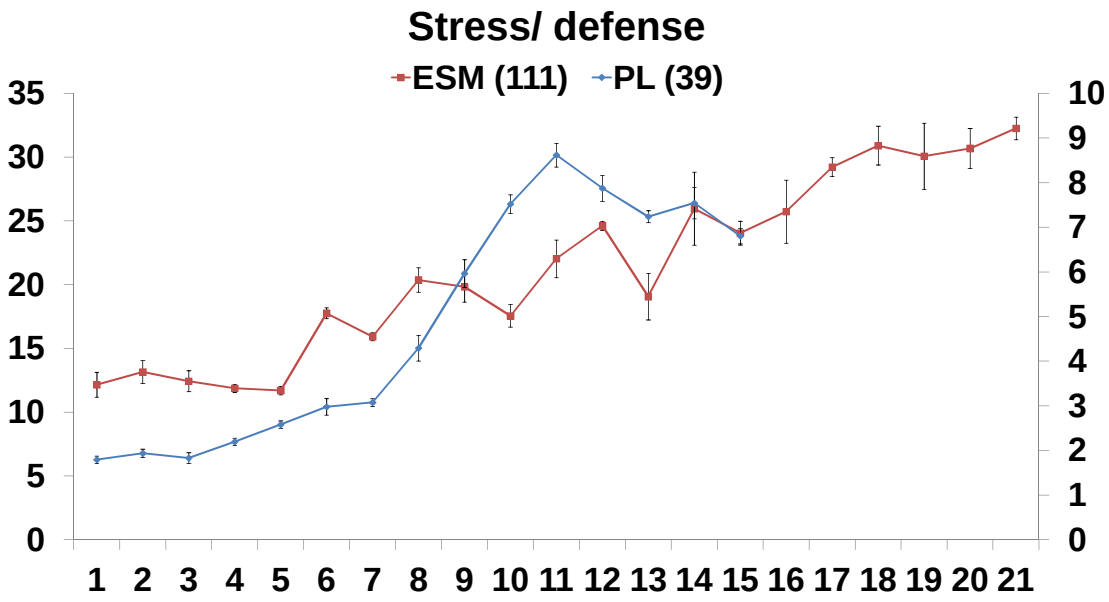
b) ATP-Interconversion (AI)

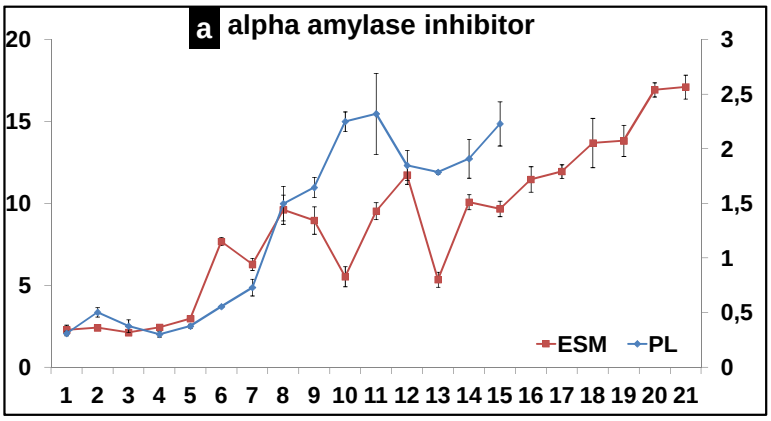
In contrast to ST and Trs categories, AI proteins peaked at the same time in PL and ESM (at stage 4). These proteins, 7 in PL and 13 in ESM, decreased after stage 4 with the same rate until stage 9 and thereafter decrease was more rapid in PL than in ESM (Figure 26 b). Different isoforms of ATP synthase and nucleoside disphosphate kinase were identified both in PL and ESM, while in ESM some additional AI related proteins were also identified.

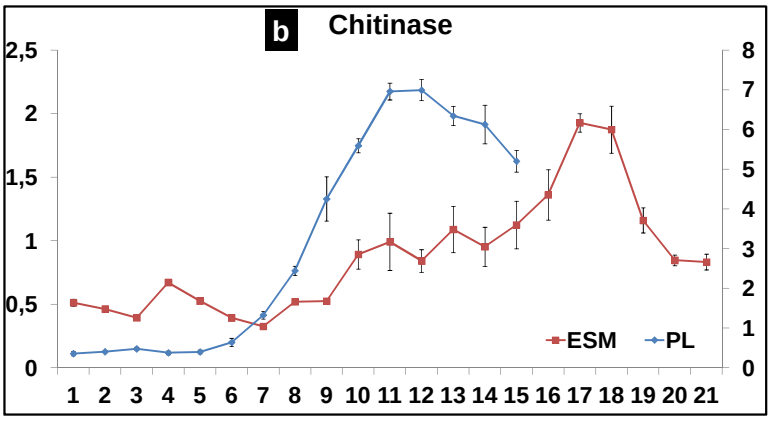
3.4 STORAGE PROTEINS (SP)

Notably in the albumin globulin fraction of ESM, we found only 1% of SP (4 proteins), while in PL, nearly 26% of proteins (51 proteins) fall into this category.

In ESM, we identified two different isoforms of globulin1, avenin-like b protein and gamma gliadin as SP proteins. Avenin-like b protein was recently studied in durum wheat and was found in gluten, leading to the hypothesis that they probably have some role in gluten functional properties (De Caro et al. 2010). Although avenin-like b and gamma gliadin are found in the gluten fraction, it is quite common to find these proteins in the metabolic fraction of starchy endosperm (Vensel et al. 2005). Storage proteins of PL were mainly globulins that, in general, are







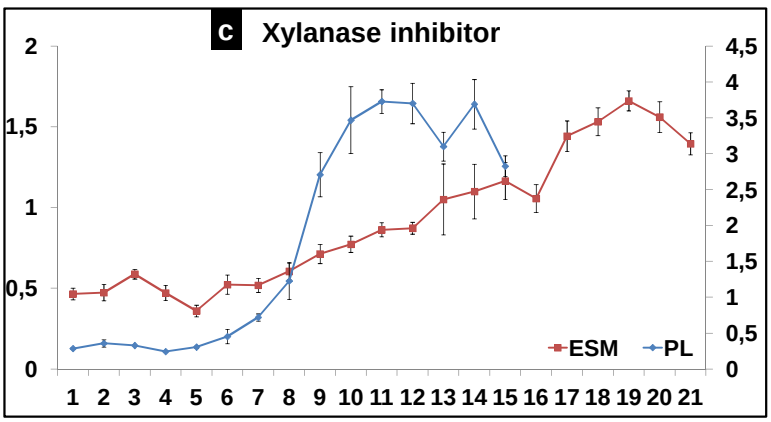


Figure 28: Protein expression profiles of stress/defense category alongwith its sub categories (**a-c**). Stages of development are on x-axis, while normalized volume values are presented on y-axis for ESM on primary axis and for PL on secondary axis.

found abundantly in aleurone layer (Halford and Shewry 2007, Loit et al. 2009). Globulin 2, 3, 3b and 3c were identified, but globulin 1 which was found in ESM, was not observed in this tissue.

These proteins in both tissues were abundant during the grain filling period up to physiological maturity, i.e stage 15 (last point of PL analysis). In ESM, the analysis was performed till stage 21, so additional information was available for this tissue, which showed decreased expression after stage 15 (Figure 27).

3.5 STRESS/ DEFENSE (SD)

Proteins related to SD category peaked towards grain maturity in both tissues (Figure 28), as observed in other studies on grain development (Vensel et al. 2005, McIntosh et al. 2007, Wan et al. 2008).

Alpha amylase inhibitor was the dominant class in ESM whereas cell wall related defense proteins (such as chitinase, xylanase inhibitor) were identified abundantly in PL. Comparison among expression profiles of these categories, in PL and ESM, revealed that, although their abundance increased towards maturity, they showed different behaviors. This can be seen for alpha amylase inhibitor, chitinase and xylanase inhibitors in Figure 28 a, b and c.

i) For alpha amylase inhibitors, a more or less linear increase with grain development was recorded.
ii) Chitinase abundance in PL, was maximum at stage 11 and then decreased, whereas for ESM they peaked at stage 17 and then decreased.
iii) Similar to chitinases, we observed a peak for xylanase inhibitors in PL (at stage 10) earlier than in ESM (at stage 19).

Other SD proteins that were identified in PL peaked at stage 15. For overall SD catgory, it is also possible that, in PL, there was again accumulation of these proteins, as they are important for grain content protection against different stresses.

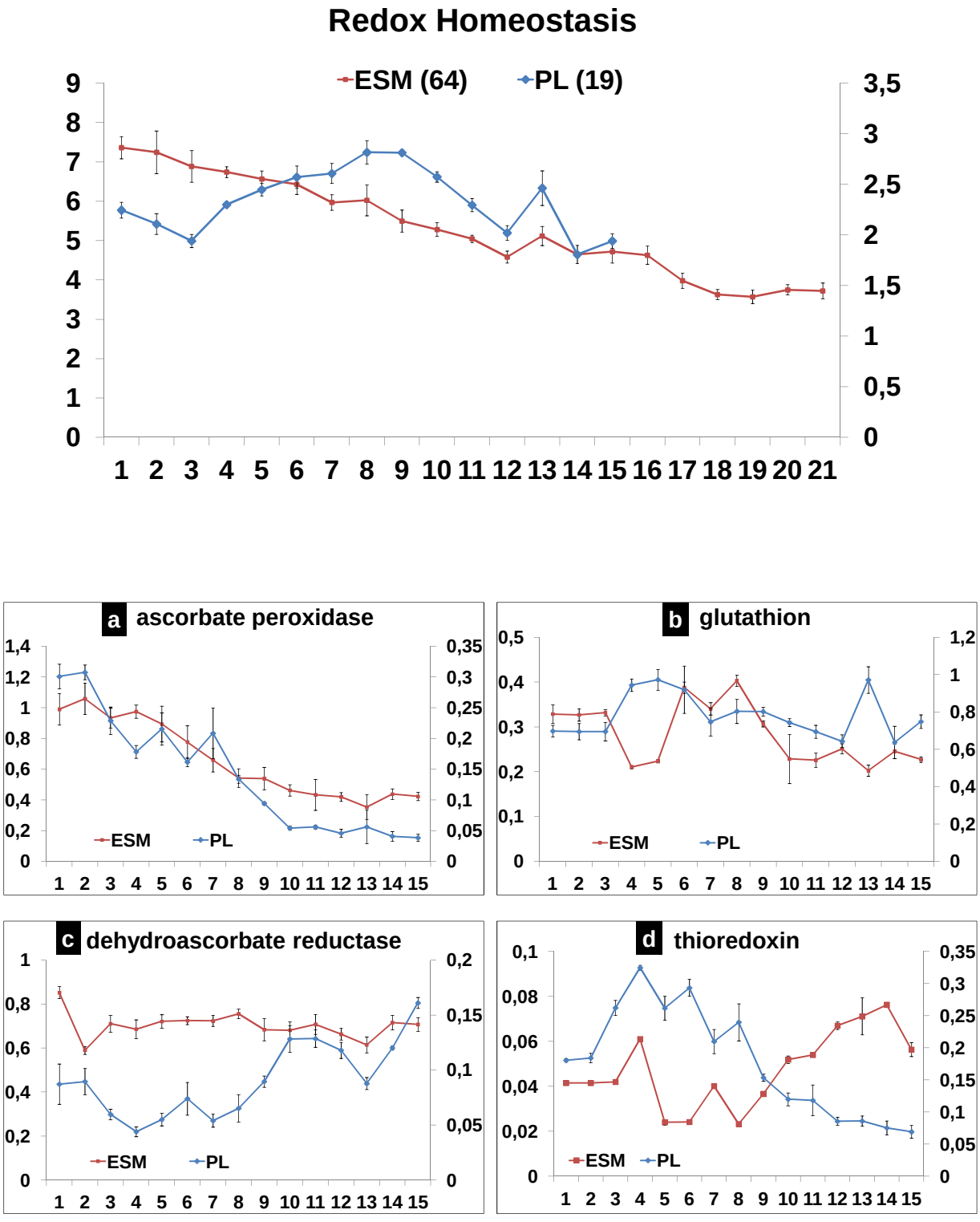


Figure 29: Redox homeostasis and its sub categories (**a-d**). Stages of development are on x-axis, while normalized volume values are presented on y-axis for ESM on primary axis and for PL on secondary axis.

3.6 REDOX HOMEOSTASIS

The process which maintains the redox environment of a cell or compartment within a cell comes under redox homeostasis. In plants, an antioxidative system determines the extent of ROS accumulation; in addition this system sets the threshold for general plant defense responses, particularly against biotic stresses and wounding. This ROS-antioxidant interaction has an important role in plant responses to environment and in the regulation of plant development (Foyer and Noctor 2005). This system is very complex and involved many agents of oxidation-reduction process. Glutathion and ascorbate are considered to play a central role in Redox system and their role in this system was recently reviewed in plants (Foyer and Noctor 2011). It was shown, using Arabidopsis mutants, that glutathion and ascorbate are multifunctional metabolites which played central roles in plant development, defense reactions and redox homeostasis & signaling (Dowdle et al. 2007, Parisy et al. 2007, Schlaeppi et al. 2008, Mhamdi et al. 2010). Kranner et al., also suggested that glutathione redox potential can act as a key determinant of cell death and dormancy (Kranner et al. 2006).

In our analysis, redox-homeostasis curves decreased towards grain maturity both in PL and ESM (Figure 29). From grain filling to grain dessication, struggle for energy and oxygen result in either a slowdown of active metabolism or switch to metabolic pathways with less ATP requirement (Huang et al. 2008). All these conditions lead to reduce metabolism both in PL and ESM (Figure 24) and also, when there is less oxygen, a slowdown of redox curve results (Figure29).

When we dissected the redox curve to investigate its ROS and antioxidant components separately, interesting results were observed. While focusing on glutathione and ascorbate activities during development, we observed that in PL and ESM, they were not abundant at the same time. In previous studies depletion of one metabolite results in activation of the other i:e when the glutahion pool is depleted, redox state control shifted to ascorbate dependent processes and vice versa (Foyer and Noctor 2011). These observations and our protein profiles (involved in redox homeostasis), helped us to develop a hypothesis concerning these metabolite activities during development. If we divide grain development from stage 1 to 15 in three phases; early, mid and late both in PL and ESM, we observed that:

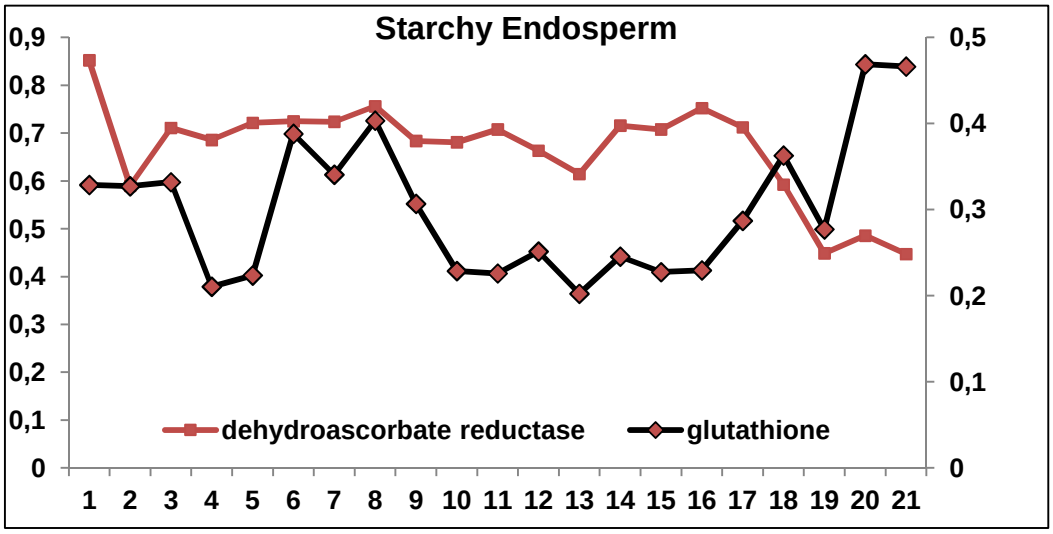


Figure 30: Proteins of Redox identified in starchy endosperm, dehydroascorbate reductase on primary and glutathione on secondary axis. Stages of development are on x-axis, while normalized volume values are presented on y-axis

- In early grain development phase, ascorbate dependent processes were active in both tissues to act as antioxidants against ROS (Figure 29 a).
- In the second phase (mid development phase), as the ascorbate pool decreased (both ascorbate and dehydroascorbate) redox control shifted to glutathione pool in PL, while in ESM, both metabolites worked in this phase interchangeably (glutathione and dehydroascorbate) (Figure 29 b & c)
- During the late development phase, in PL, the control shifted again to the ascorbate pool (now dehydroascorbate), also in ESM glutathione control was reduced considerably and the ascorbate pool become dominant (dehydroascorbate) (Figure 29 c).

ESM data, after stage 15, revealed additional information that after this stage, the Redox environment control again shifted from the ascorbate pool to the glutathione pool (Figure 30).

Thioredoxin curves showed opposite behavior in the two tissues, they peaked during early stages in PL and during late development stages in ESM (Figure 29 d). In PL, the early presence would help catabolism of cross and tube cells, while in ESM, due to reorganization of storage proteins, they appear at later stages. In addition, this may be part of grain preparation process for germination, after which thioredoxin help for catabolism of storage proteins (Li et al. 2009).

These data of redox homeostasis are based on protein identifications; further enzymatic studies would be helpful to improve understanding and to validate our hypotheses. Moreover, comparison with a third grain tissue, the embryo, particularly at germination stages and further during development would provide convincing data of developmental changes. A preliminary example of such comparison is provided by Wan et al., by using transcriptional profiling of whole grain, and then putative tissue assignation using data published on barley by Sreenivasulu et al., 2006 and on wheat by Drea et al., 2005.

CONCLUSION AND PERSPECTIVES

This study explored the changes during development of two important parts of wheat grain: nutritionally important peripheral layers and yield determining starchy endosperm. Further, a first attempt to link proteome and transcriptome data in wheat has been provided in this study. The major results were:

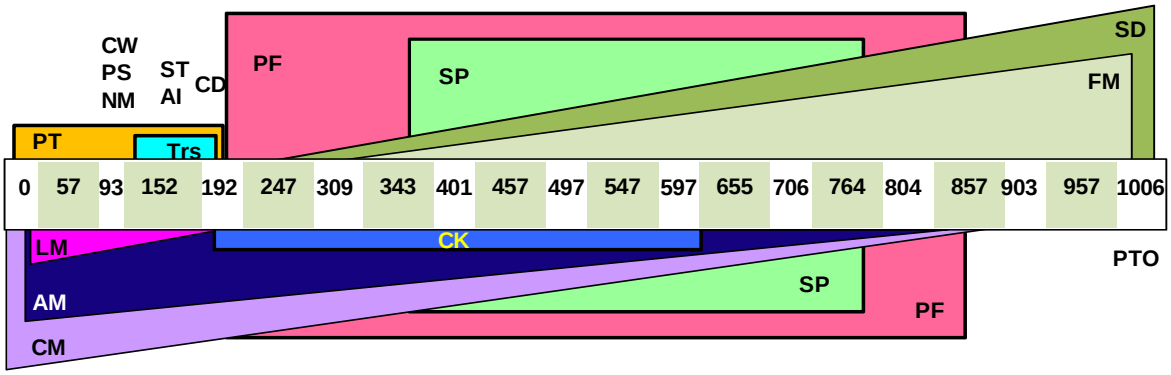
- For the first time in wheat, we provided proteomic data for grain peripheral layer which included 207 proteins identified during development: from fertilization to physiological maturity of grain at fifteen stages of development.
- This study revealed five different expression profiles during development.
- Proteins were classed in 16 different functional categories. If we divide complete development into early, mid and late, the following processes were abundant at each of these three stages.

Early	**Mid**	**Late**
Metabolism	Protein synthesis	
Translation	Protein turn over	Storage Proteins
Transcription	signal transduction	Stress / defense
ATP interconversion	membrane transport	
	Biosynthesis of secondary metabolites	

Similarly for starchy endosperm metabolic proteins, we developed a proteome map at twenty one stages of development: from fertilization to complete grain maturity. The important results were:

- Identification of 487 proteins differentially expressed during development.
- Nine different expression profiles were evident during development.
- Prediction of sub-cellular localizations for all identified proteins.
- Dissection of SD category in its components and their analysis.
- Proteins were classed in 17 different biochemical functional classes and these functions differ among expression profiles. As for PL data, to obtain an overview of functions during development, if we divide grain development into three phases

a: Starchy endosperm



b: Peripheral Layers

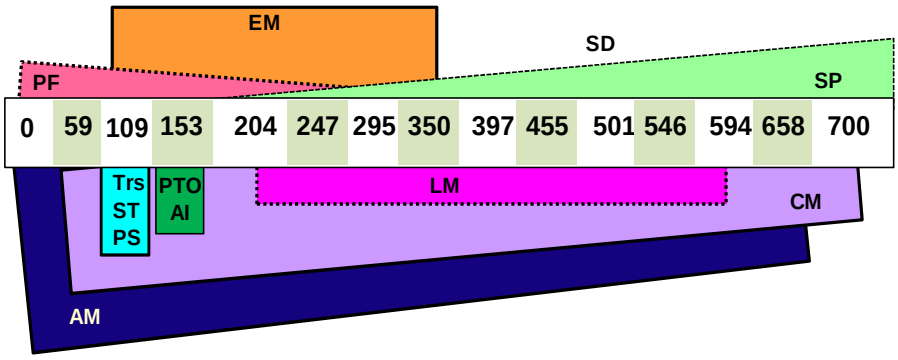


Figure 31: Schematic presentation of functional categories during development of two kernel tissues **a)** starchy endosperm **b)** peripheral layers.

(early, mid and late) these functions were distributed according to their abundance as follow:

Early and Early to Mid	**Mid**	**Late**
Metabolism		
Cell division	Protein folding	
Protein transport	Storage proteins	Stress/ defense
ATP inter-conversion	Proteins related	Folic acid metabolism
Protein synthesis	to cytoskeleton	Protein turn over
Transcription/translation		
Signal transduction		

A development study with very short temporal distances, enabled us to determine precisely the stage or stages of development in which each category appeared abundantly, as can be seen in Figure 4 of Tasleem-Tahir et al., 2011 and Figure 3 of Tasleem-Tahir et al., 2012.

We discuss the differences and similarities which we observed in the proteomic analyses, of peripheral layers and starchy endosperm during development. This comparison is schematized in Figure 31, and led to the following conclusions:

- Changes during grain development of CM, AM, PS, Trs, AI, ST and Redox related proteins were similar in both tissues
- LM, PF and PTO showed different behaviors in PL and ESM
- For SP and SD, up to 700°Cd, changes were quite similar, but in addition for ESM after this stage, we observed an increase of expression.
- For Redox homeostasis during development of tissues, glutathione and ascorbate pools control ROS accumulation alternately in the two tissues i:e during development in one tissue when there was depletion of one pool the other one controlled homeostasis. In addition it was noticed that their abundance was not same in both tissues at a given stage.

Finally we made the first attempt to integrate proteome, transcriptome data for wheat, using the 145 proteins which were identified for carbohydrate metabolism in starchy endosperm.

- Nearly 32% of pairs showed significant correlations, of which 6% were correlated negatively and 26% had positive significant correlations.
- We divide this metabolism into five main categories, and their enzymes at proteome and transcript level were mapped using Mapman in CM proposed pathway. These categories were, sucrose & starch category; fructose, mannose & galactose; pentose phosphate metabolism; glycolysis; and citrate cycle or TCA.
- Among these five categories, we showed the highest significant correlations for starch and sucrose category (43%), between proteome and transcriptome data

This study leads to many avenues for future research, which could contribute to grain quality improvement.

Wheat grain peripheral layers are important in several ways, and each layer has its own characteristics and participation in grain development, nutritional quality and health value. During grain development, studies of individual layers should be made: such as aleurone layer, for health and nutritional benefits. In addition, for study of phytohormone signaling in germination-related processes, aleurone layer could be used as model system as was recently reviewed in barley (Finnie et al. 2011).

Cross and tube cells should be studied for their photosynthetic activities and response to hormones (such as auxins involved for cell division and elongation) during formation of lignified structures which determine the size and dimensions of grain.

Study of testa during development could be helpful to understand its role as a transfer tissue which operates between pericarp and aleurone layer. These outer layers together form a gateway for any pathogen or external attack. Each layer has its distinct line of defense for embryo and nutrient-rich endosperm protection (Jerkovic et al. 2010). Comparative studies between normal and diseased grains using only PL part may provide important clues regarding grain protection against these external unwanted factors. For example, Fusarium Head Blight is world wide problem of wheat, decreasing yield and quality. Mycotoxins produced by this fungus are harmful when wheat products are used for food by humans and animals (Bennett and Klich 2003). Study

of peripheral layers during kernel development in plants with and without attack could provide important information about mechanisms of disease development. This would be an important tool in research for disease remedies.

Identification of different isoforms of the same protein is an indication of probable post translational modification occurrence. A further analysis is needed to verify these PTM either by use of MS or other techniques for functional analysis.

In depth analysis to reveal the proteins that are less abundant is possible using more sensitive spot detection methods and MS tools. Similarly, the gel free MALDI imaging methods could also be a help to study the proteins in different grain tissues (such as aleurone layer, endosperm, and embryo). At present their applications in plants are very limited but in the near future, they could be useful in plant research for rapid investigations (Kaspar et al. 2011, Lee et al. 2012).

Preliminary comparison of proteome and transcriptome requires further data enrichment and verification. For example PCR analysis to check expression profiles of genes would be helpful. Further, if we broaden our analysis by using other functional categories in addition to CM proteins, this could provide additional annotation for transcript data and moreover additional information about their sequences.

To complement the metabolic proteins, we also performed 2D and HPLC analysis of gluten proteins during development. This data is in the course of analysis, the comparison of expression profiles of metabolic proteins identified in ESM and that of gluten proteins would help not only to improve understanding of the changes in these grain reserve proteins but also the influence of metabolic proteins on their accumulation.

Similarly, application of proteome/transcriptome comparison could provide additional information about this gluten complex during development. Together, all these information would be helpful for creation of a Récital databank, which would allow us to understand the mechanisms involved in grain formation and could be used as a reference for comparative studies of different biotic and abiotic stresses.

In near future genetic analyses of wheat nutritional and health value will be based on proteomic and transcriptomic data to provide information regarding stage and tissues specific functional markers.

Modeling grain growth and size and its constituents is becoming popular area of research (Martre et al. 2011). Combining the data of agro-physiologists and that of proteome and transcriptome during development would be helpful for developing models of accumulation of grain constituents.

REFERENCES

REFERENCES

Acevedo E., P. Silva, H. Silva. 2006. Growth and Wheat Physiology, Development. Laboratory of Soil Plant Water Relations. Faculty of Agronomy and Forestry Sciences University of Chile, Chile, pp 1–47.

Adams, M., J. Kelley, J. Gocayne, M. Dubnick, M. Polymeropoulos, H. Xiao, C. Merril, A. Wu, B. Olde, R. Moreno, and a. et. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252: 1651-1656.

Aebersold, R., and M. Mann. 2003. Mass spectrometry-based proteomics. Nature 422: 198-207.

Altenbach, S.B., F.M. DuPont, K.M. Kothari, R.Chan, E.L. Johnson, D. Lieu. 2003. Temperature, water and fertilizer influence the timing of key events during grain development in US spring wheat. Journal of Cereal Science 37: 9–20.

Antoine, C., V. Lullien-Pellerin, J. Abecassis, and X. Rouau. 2002. Nutritional interest of the wheat seed ale rone layer. Sciences Des Aliments 22: 545-556.

Antoine, C., S. Peyron, F. Mabille, C. Lapierre, B. Bouchet, J. Abecassis, and X. Rouau. 2003. Individual contribution of grain outer layers and their cell wall structure to the mechanical properties of wheat bran. Journal of Agricultural and Food Chemistry 51: 2026-2033.

Aoki, N., G. N. Scofield, X.-D. Wang, C. E. Offler, J. W. Patrick, and R. T. Furbank. 2006. Pathway of Sugar Transport in Germinating Wheat Seeds. Plant physiology 141: 1255-1263.

Baker, C.K., and J.N. Gallagher. 1983a. The development of winter wheat in the field. Relation between apical development and plant morphology within and between seasons. J Agric Sci Cambridge, 10: 327-335.

Baker, C.K., and J.N. Gallagher. 1983b. The development of winter wheat in the field. The control of primordium initiation rate by temperature and photoperiod. J Agric Sci Cambridge, 101: 337-344.

Baldo, B., A. Reid, and P. Boniface. 1983. Lectins as Cytochemical Probes of the Developing Wheat Grain. IV. Demonstration of Mucilage Containing L-Fucose Associated With Roots in Ungerminated Grain. Functional Plant Biology 10: 459-470.

Barron, C., A. Surget, and X. Rouau. 2007. Relative amounts of tissues in mature wheat (*Triticum aestivum* L.) grain and their carbohydrate and phenolic acid composition. Journal of Cereal Science 45: 88-96.

Belderok, B. 2000. Developments in bread-making processes. Plant foods for human nutrition 55: 1-86.

Bennett, M. D., J. B. Smith, and I. Barclay. 1975. Early Seed Development in the Triticeae. Philosophical Transactions of the Royal Society of London. B, Biological Sciences 272: 199-227.

Bennett, J. W., and M. Klich. 2003. Mycotoxins. Clinical microbiology reviews 16: 497-516.

Bietz, J.A, J.S. Wall. 1972. Wheat gluten subunits: molecular weights determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis. Cereal Chemistry, 49: 416-430.

Bietz, J.A., J.S. Wall. 1980. Identity of high molecular weight gliadin and ethanol-soluble glutenin subunits of wheat: relation to gluten structure. Cereal Chemistry, 57:415-421.

Bindschedler, L. V., M. Palmblad, and R. Cramer. 2008. Hydroponic isotope labelling of entire plants (HILEP) for quantitative plant proteomics; an oxidative stress case study. Phytochemistry 69: 1962-1972.

Bindschedler, L. V., and R. Cramer. 2011. Quantitative plant proteomics. Proteomics 11: 756-775.

Blackstock, W. P., and M. P. Weir. 1999. Proteomics: quantitative and physical mapping of cellular proteins. Trends in Biotechnology 17: 121-127.

Bradbury, D., M.M. MacMasters, and I.M Cull. 1956a. Structure of the mature wheat kernel. II. Microscopic Structure of pericarp, seed coat, and other coverings of the endosperm and germ of hard red winter wheat. Cereal Chemistry 33: 342-346.

Bradbury, D., M.M. MacMasters, and I.M. Cull. 1956b. Structure of the mature wheat kernel. III. Microscopic Structure of the endosperm of hard red winter wheat. Cereal Chemistry 33: 361-373.

Branlard G, D. Marion. 2011. Proteins and lipids for 21st century. In: Bonjean A, Angus, W (Ed) World Wheat Book II, in preparation.

Brocklehurst, P. A. 1977. Factors controlling grain weight in wheat. Nature 266: 348-349.

Brouns, F., Y. Hemery, R. Price, and N. M. Anson. 2012. Wheat aleurone: separation, composition, health aspects, and potential food use. Critical reviews in food science and nutrition 52: 553-568.

Campbell, W., J. Lee, T. O'brien, and M. Smart. 1981. Endosperm Morphology and Protein Body Formation in Developing Wheat Grain. Functional Plant Biology 8: 5-19.

Chen, W.J., X. Fan, B. Zhang, B.L. Liu, Z.H. Yan, L.Q. Zhang, Z.W. Yuan, Y.L. Zheng, H.G. Zhang, and D.C. Liu. 2012. Novel and ancient HMW glutenin genes from Aegilops tauschii and their phylogenetic positions. Genetic Resources and Crop Evolution.

Chu, C.G., C.T. Tan, G.T. Yu, S. Zhong, S.S. Xu, and L. Yan. 2011. A Novel Retrotransposon Inserted in the Dominant Vrn-B1 Allele Confers Spring Growth Habit in Tetraploid Wheat (*Triticum turgidum* L.). G3Journal 1: 637-645.

Clare Mills, E. N., S. R. Burgess, A. S. Tatham, P. R. Shewry, H. W. S. Chan, and M. R. A. Morgan. 1990. Characterization of a panel of monoclonal anti-gliadin antibodies. Journal of Cereal Science 11: 89-101.

Coram, T . E ., G. Brown-Guedira and X . Chen. 2008. Using transcriptomics to understand the wheat genome. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources 083 : 1-9

Cornell, H., 2003. In: Cauvain SP (ed) Bread Making: Improving Quality. Woodhead Publishing, Cambridge.

Davidson, R. M., M. Gowda, G. Moghe, H. Lin, B. Vaillancourt, S. H. Shiu, N. Jiang, and C. Robin Buell. 2012. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. The Plant journal : for cell and molecular biology.

De Caro, S., P. Ferranti, F. Addeo, and G. Mamone. 2010. Isolation and characterization of Avenin-like protein type-B from durum wheat. Journal of Cereal Science 52: 426-431.

Dennis, E. S., and W. J. Peacock. 2009. Vernalization in cereals. Journal of biology 8: 57.

Domon, B., and R. Aebersold. 2006. Mass Spectrometry and Protein Analysis. Science 312: 212-217.

Dorofeev, V.F., A.A. Filatenko, E.F. Migushova, R.A. Udaczin, and M.M. Jakubziner. 1979. Wheat. vol. 1. In: *Flora of Cultivated Plants* (Dorofeev VF and Korovina ON, Eds.) Leningrad (St. Petersburg), Russia. Kolos (in Russian). 346 pp.

Dowdle, J., T. Ishikawa, S. Gatzek, S. Rolinski, and N. Smirnoff. 2007. Two genes in Arabidopsis thaliana encoding GDP-l-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. The Plant Journal 52: 673-689.

Drea, S., D. J. Leader, B. C. Arnold, P. Shaw, L. Dolan, and J. H. Doonan. 2005. Systematic spatial analysis of gene expression during wheat caryopsis development. The Plant cell 17: 2172-2185.

Dreger, M. 2003. Proteome analysis at the level of subcellular structures. European Journal of Biochemistry 270: 589-599.

Dubcovsky, J., and J. Dvorak. 2007. Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. Science 316: 1862-1866.

Dvorak, J., P. Di Terlizzi, H. B. Zhang and P. Resta A. 1993. The evolution of polyploid wheats: identification of the A genome donor species. Genome 36: 21-31.

Encinas, P., M. A. Rodriguez-Milla, B. Novoa, A. Estepa, A. Figueras, and J. Coll. 2010. Zebrafish fin immune responses during high mortality infections with viral haemorrhagic septicemia rhabdovirus. A proteomic and transcriptomic approach. BMC Genomics 11.

Engelsberger, W. R., A. Erban, J. Kopka, and W. X. Schulze. 2006. Metabolic labeling of plant cell cultures with K(15)NO3 as a tool for quantitative analysis of proteins and metabolites. Plant methods 2: 14.

Esau, K. 1953. Plant anatomy. London, Chapman & Hall.

Evans, L.T., I.F. Wardlaw, R.A. Fischer. 1975. Wheat. In L.T. Evans, eds. Crop Physiology Cambridge University Press. USA, pp 101-149.

Evers, A. D. 1970. Development of the Endosperm of Wheat. Annals of Botany 34: 547-555.
Evers, T., and S. Millar. 2002. Cereal Grain Structure and Development: Some Implications for Quality. Journal of Cereal Science 36: 261-284.
Feuillet, C., P. Langridge, and R. Waugh. 2008. Cereal breeding takes a walk on the wild side. Trends in genetics : TIG 24: 24-32.
Feldman, M. 1995. Wheats. *In:* Smartt J, Simmonds NW, *1995 editors.* Evolution of crop plants. Harlow, UK*:* Longman Scientific and Technical, *pp* 185-192.
Feldman, M. 2001. Origin of cultivated wheat. *In:* Bonjean AP, Angus WJ, *editors.* The world wheat book: a history of wheat breeding. Paris, France*:* Lavoisier Publishing, *pp* 3-56.
Ferreira, M.S.L., P. Martre, C. Mangavel, C. Girousse, N. N. Rosa, M.-F. Samson, and M.-H. Morel. 2011. Physicochemical control of durum wheat grain filling and glutenin polymer assembly under different temperature regimes. Journal of Cereal Science (article in press).
Fincher, G.B. and B.A. Stone. 1986. Cell walls and their components in cereal grain technology. *In*: "Advances in Cereal Science and Technology volume 8" (Pomeranz, Y. ed.) Am. Assoc. Cereal Chem., St. Paul, Minnesota, pp 207–295.
Finnie, C., and B. Svensson. 2003. Feasibility study of a tissue-specific approach to barley proteome analysis: aleurone layer, endosperm, embryo and single seeds. Journal of Cereal Science 38: 217-227.
Finnie, C., B. Andersen, A. Shahpiri, and B. Svensson. 2011. Proteomes of the barley aleurone layer: A model system for plant signalling and protein secretion. Proteomics 11: 1595-1605.
Flood, R.G., and G.M. Halloran. 1986. Genetics and physiology of vernalisation response in wheat. Adv. Agron., 39: 87-125.
Fossati, D., and M. Ingold. 2001. Mountain wheat pool. *In:* Bonjean AP, Angus WJ, *editors.* The world wheat book: a history of wheat breeding. Paris, France*:* Lavoisier Publishing, *pp* 311-332.
Forne, I., B. Castellana, R. Marin-Juez, J. Cerda, J. Abian, and J. V. Planas. 2011. Transcriptional and proteomic profiling of flatfish (Solea senegalensis) spermatogenesis. Proteomics 11: 2195-2211.
Foyer, C. H., and G. Noctor. 2005. Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. The Plant cell 17: 1866-1875.
Foyer, C. H., and G. Noctor. 2011. Ascorbate and Glutathione: The Heart of the Redox Hub. Plant physiology 155: 2-18.
Gale, K. R. 2005. Diagnostic DNA markers for quality traits in wheat. Journal of Cereal Science 41: 181-192.
Gallagher, J.N. and P.V. Biscoe. 1978. A physiological analysis of cereal yield. II. Partitioning of dry matter. Agric Prog 53: 51-70.
Gallardo, K., C. Firnhaber, H. Zuber, D. Hericher, M. Belghazi, C. Henry, H. Kuster, and R. Thompson. 2007. A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds. Molecular and Cellular Proteomics 6: 2165-2179.
Gardner, J.S., W.M Hess., and E. J. Trione. 1985. Development of the young wheat spike: a SEM study of Chinese spring wheat. *American Journal of Botany* 72: 548-559.
Ghazalpour, A., B. Bennett, V. A. Petyuk, L. Orozco, R. Hagopian, I. N. Mungrue, C. R. Farber, J. Sinsheimer, H. M. Kang, N. Furlotte, C. C. Park, P. Z. Wen, H. Brewer, K. Weitz, D. G. Camp, C. Pan, R. Yordanova, I. Neuhaus, C. Tilford, N. Siemers, P. Gargalovic, E. Eskin, T. Kirchgessner, D. J. Smith, R. D. Smith, and A. J. Lusis. 2011. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. Plos Genetics 7.
Gorg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss. 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21: 1037-1053.
Görg, A., W. Postel, and S. Günther. 1988. Two-dimensional electrophoresis. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 9: 531-546.

Görg, A., W. Weiss, and M. J. Dunn. 2004. Current two-dimensional electrophoresis technology for proteomics. Proteomics 4: 3665-3685.

Greer, E.N., J.J.C. Hinton, C.R. Jones, and N.L. Kent. 1951. The occurrences of endosperm cells in wheat flour. Cereal Chemistry 28: 58-67.

Groos, C., E. Bervas, and G. Charmet. 2004. Genetic analysis of grain protein content, grain hardness and dough rheology in a hard×hard bread wheat progeny. Journal of Cereal Science 40: 93-100.

Gruhler, A., W. X. Schulze, R. Matthiesen, M. Mann, and O. N. Jensen. 2005. Stable Isotope Labeling of Arabidopsis thaliana Cells and Quantitative Proteomics by Mass Spectrometry. Molecular and Cellular Proteomics 4: 1697-1709.

Gupta, R. B., K. W. Shepherd, and F. MacRitchie. 1991. Genetic control and biochemical properties of some high molecular weight albumins in bread wheat. Journal of Cereal Science 13: 221-235.

Gupta, R.B., and F. MacRitchie. 1994. Allelic variation at glutenin subunit and gliadin loci, *Glu-1-Glu-3* and *Gli-1,* of common wheats. II. Biochemical basis of the allelic effects on dough properties, Journal of Cereal Science, 19: 19–29.

Gustavsson, N., B. Greber, T. Kreitler, H. Himmelbauer, H. Lehrach, and J. Gobom. 2005. A proteomic method for the analysis of changes in protein concentrations in response to systemic perturbations using metabolic incorporation of stable isotopes and mass spectrometry. Proteomics 5: 3563-3570.

Hajduch, M., L. B. Hearne, J. A. Miernyk, J. E. Casteel, T. Joshi, G. K. Agrawal, Z. Song, M. Zhou, D. Xu, and J. J. Thelen. 2010. Systems analysis of seed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression. Plant physiology 152: 2078-2087.

Halford, N., and P. Shewry. 2007. The Structure and Expression of Cereal Storage Protein Genes Endosperm, pp. 195-218. In O.-A. Olsen (ed.), vol. 8. Springer Berlin / Heidelberg.

Halford, N. G., J. M. Field, H. Blair, P. Urwin, K. Moore, L. Robert, R. Thompson, R. B. Flavell, A. S. Tatham, and P. R. Shewry. 1992. Analysis of HMW glutenin subunits encoded by chromosome 1A of bread wheat (*Triticum aestivum* L.) indicates quantitative effects on grain quality. TAG Theoretical and Applied Genetics 83: 373-378.

Hanft, J. M., and R. D. Wych.1982. Visual Indicators of Physiological Maturity of Hard Red Spring Wheat. Crop Science 22: 584-588.

Hay, R.K.M., and E.J.M. Kirby. 1991. Convergence and synchrony-A review of the coordination of development in wheat. Aust J Agric Res 42: 661-700.

Heun, M., R. Schäfer-Pregl, D. Klawan, R. Castagna, M. Accerbi, B. Borghi, F. Salamini. 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278:1312-1314.

Huang, S., T. D. Colmer, and A. H. Millar. 2008. Does anoxia tolerance involve altering the energy currency towards PPi? Trends in plant science 13: 221-227.

Huillet, C., A. Adrait, D. Lebert, G. Picard, M. Trauchessec, M. Louwagie, A. Dupuis, L. Hittinger, B. Ghaleb, P. Le Corvoisier, M. Jaquinod, J. Garin, C. Bruley, and V. Brun. 2012. Accurate Quantification of Cardiovascular Biomarkers in Serum Using Protein Standard Absolute Quantification (PSAQ™) and Selected Reaction Monitoring. Molecular and Cellular Proteomics 11.

Jerkovic, A., A. M. Kriegel, J. R. Bradner, B. J. Atwell, T. H. Roberts, and R. D. Willows. 2010. Strategic Distribution of Protective Proteins within Bran Layers of Wheat Protects the Nutrient-Rich Endosperm. Plant physiology 152: 1459-1470.

Jestin, L., C. Ravel, S. Auroy, B. Laubin, M. R. Perretant, C. Pont, and G. Charmet. 2008. Inheritance of the number and thickness of cell layers in barley aleurone tissue (*Hordeum vulgare* L.): an approach using F2-F3 progeny. TAG. Theoretical and applied genetics 116: 991-1002.

Johnson, B.L. and H.S. Dhaliwal. 1976. Reproductive isolation of *Triticum boeoticum* and *Triticum urarta* and the origin of the tetraploid wheat. American Journal of Botany 63: 1088-1094.

Johnson, J. C., R. Appels, and M. Bhave. 2006. The PDI genes of wheat and their syntenic relationship to the esp2 locus of rice. Funct Integr Genomics 6: 104-121.

Kamal-Eldin, A., H.N. Lærke, K.E. Bach Knudsen, A.M. Lampi, V. Piironen, H. Adlercreutz *et al,* 2009. Physical, microscopic, and chemical characterisation of industrial rye and wheat brans from the Nordic countries, Food and Nutrition Research 53.

Kaspar, S., M. Peukert, A. Svatos, A. Matros, and H.-P. Mock. 2011. MALDI-imaging mass spectrometry – An emerging technique in plant biology. PROTEOMICS 11: 1840-1850.

Kawaura, K., K. Mochida, and Y. Ogihara. 2005. Expression profile of two storage-protein gene families in hexaploid wheat revealed by large-scale analysis of expressed sequence tags. Plant physiology 139: 1870-1880.

Kihara, H., 1944. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare* (Japenese). Agric Hort 19: 13-14.

Kirby, E.J.M. and M. Appleyard. 1984. Cereal development guide. Arable Unit, National Agriculture Centre, Stoneleigh, Kenilworth. England, pp 95.

Kolster, P., F. A. Eeuwijk, and W. M. J. Gelder. 1991. Additive and epistatic effects of allelic variation at the high molecular weight glutenin subunit loci in determining the bread-making quality of breeding lines of wheat. Euphytica 55: 277-285.

Koubaa, M., S. Mghaieth, B. Thomasset, and A. Roscher. 2012. Gas chromatography-mass spectrometry analysis of (13)C labeling in sugars for metabolic flux analysis. Analytical biochemistry.

Kranner, I., S. Birtić, K. M. Anderson, and H. W. Pritchard. 2006. Glutathione half-cell reduction potential: A universal stress marker and modulator of programmed cell death? Free Radical Biology and Medicine 40: 2155-2165.

Krijgsveld, J., R. F. Ketting, T. Mahmoudi, J. Johansen, M. Artal-Sanz, C. P. Verrijzer, R. H. A. Plasterk, and A. J. R. Heck. 2003. Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. Nat Biotech 21: 927-931.

Krumm, M., V. Moazami, and P. Martin. 1990. Influence of potassium nutrition on concentrations of water soluble carbohydrates, potassium, calcium, and magnesium and the osmotic potential in sap extracted from wheat (*Triticum aestivum*) ears during preanthesis development. Plant and Soil 124: 281-285.

Kumar, P., R.K. Yadava, B.Gollen, S. Kumar, R.K. Verma, S. Yadav. 2011. Nutritional Contents and Medicinal Properties of Wheat: A Review. Life Sciences and Medicine Research 22: 1-10.

Lafiandra, D., R. D'Ovidio, E. Porceddu, B. Margiotta, G. Colaprico. 1993. New data supporting high Mr glutenin subunit 5 as determinant of quality differences among the pairs 5+10 vs 2+12, Journal of Cereal Science 18: 197–205.

Lamacchia, C., P. R. Shewry, N. Di Fonzo, J. L. Forsyth, N. Harris, P. A. Lazzeri, J. A. Napier, N. G. Halford, and P. Barcelo. 2001. Endosperm-specific activity of a storage protein gene promoter in transgenic wheat seed. J Exp Bot 52: 243-250.

Large, E. C. 1954. Growth stages in Cereals illustration of the feekes scale. Plant Pathology 3: 128-129.

Lau, A. T., Q. Y. He, and J. F. Chiu. 2003. Proteomic technology and its biomedical applications. Sheng wu hua xue yu sheng wu wu li xue bao Acta biochimica et biophysica Sinica 35: 965-975.

Laudencia-Chingcuanco, D. L., B. S. Stamova, G. R. Lazo, X. Cui, and O. D. Anderson. 2006. Analysis of the wheat endosperm transcriptome. Journal of applied genetics 47: 287-302.

Laudencia-Chingcuanco, D. L., B. S. Stamova, F. M. You, G. R. Lazo, D. M. Beckles, and O. D. Anderson. 2007. Transcriptional profiling of wheat caryopsis development using cDNA microarrays. Plant Mol Biol 63: 651-668.

Law, C.N., A.J. Worland, B. Giorgi. 1975. The genetic control of ear emergence time by chromosomes 5A and 5D of wheat. Heredity 36: 49–58.

Le Trionnaire, G., F. Francis, S. Jaubert-Possamai, J. Bonhomme, E. De Pauw, J. P. Gauthier, E. Haubruge, F. Legeai, N. Prunier-Leterme, J. C. Simon, S. Tanguy, and D. Tagu. 2009. Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. BMC genomics 10: 456.

Leader, D. J. 2005. Transcriptional analysis and functional genomics in wheat. Journal of Cereal Science 41: 149-163.

Lee, Y. J., D. C. Perdian, Z. Song, E. S. Yeung, and B. J. Nikolau. 2012. Use of mass spectrometry for imaging metabolites in plants. The Plant Journal 70: 81-95.

Li, C. Y., C. Li, Z. X. Lu, W. H. Li, and L. P. Cao. 2012. Morphological changes of starch granules during grain filling and seed germination in wheat. Starch - Stärke 64: 166-170.

Li, Y.-C., J.-P. Ren, M.-J. Cho, S.-M. Zhou, Y.-B. Kim, H.-X. Guo, J.H. Wong, H.-B. Niu, H.-K. Kim, S. Morigasaki, P.G. Lemaux, O.L. Frick, J. Yin, and B.B. Buchanan. 2009. The Level of Expression of Thioredoxin is Linked to Fundamental Properties and Applications of Wheat Seeds. *Molecular Plant*. 2:430-441.

Liang, P., and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257: 967-971.

Linkiewicz, A. M., L. L. Qi, B. S. Gill, A. Ratnasiri, B. Echalier, S. Chao, G. R. Lazo, D. D. Hummel, O. D. Anderson, E. D. Akhunov, J. Dvorak, M. S. Pathan, H. T. Nguyen, J. H. Peng, N. L. Lapitan, Miftahudin, J. P. Gustafson, C. M. La Rota, M. E. Sorrells, K. G. Hossain, V. Kalavacharla, S. F. Kianian, D. Sandhu, S. N. Bondareva, K. S. Gill, E. J. Conley, J. A. Anderson, R. D. Fenton, T. J. Close, P. E. McGuire, C. O. Qualset, and J. Dubcovsky. 2004. A 2500-locus bin map of wheat homoeologous group 5 provides insights on gene distribution and colinearity with rice. Genetics 168: 665-676.

Liu, Y., N. Ye, R. Liu, M. Chen, and J. Zhang. 2010. H2O2 mediates the regulation of ABA catabolism and GA biosynthesis in Arabidopsis seed dormancy and germination. *Journal of Experimental Botany*. 61:2979-2990.

Lockhart, D. J., H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Norton, and E. L. Brown. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotech 14: 1675-1680.

Loit, E., C. Melnyk, A. MacFarlane, F. Scott, and I. Altosaar. 2009. Identification of three wheat globulin genes by screening a Triticum aestivum BAC genomic library with cDNA from a diabetes-associated globulin. BMC plant biology 9: 93.

Lopes, M. A., and B. A. Larkins. 1993. Endosperm origin, development, and function. The Plant cell 5: 1383-1399.

Lottspeich, F., and J. Kellermann. 2011. ICPL labeling strategies for proteome research. Methods in molecular biology (Clifton, N.J.) 753: 55-64.

Loussert, C., Y. Popineau, and C. Mangavel. 2008. Protein bodies ontogeny and localization of prolamin components in the developing endosperm of wheat caryopses. Journal of Cereal Science 47: 445-456.

Lukow, O. M., P. I. Payne, and R. Tkachuk. 1989. The HMW glutenin subunit composition of Canadian wheat cultivars and their association with bread-making quality. Journal of the Science of Food and Agriculture 46: 451-460.

Luo, C., W. B. Griffin, G. Branlard, and D. L. McNeil. 2001. Comparison of low- and high molecular-weight wheat glutenin allele effects on flour quality. Theoretical and Applied Genetics 102: 1088-1098.

Lutz, E., H. Wieser, and P. Koehler. 2012. Identification of Disulfide Bonds in Wheat Gluten Proteins by Means of Mass Spectrometry/Electron Transfer Dissociation. Journal of Agricultural and Food Chemistry 60: 3708-3716.

Macritchie, F. 1992. Physicochemical Properties of Wheat Proteins in Relation to Functionality, pp. 1-87. In E. K. John (ed.), Advances in Food and Nutrition Research, vol. Volume 36. Academic Press.

Mann, M. 2006. Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol 7: 952-958.

Martin, J.H., W.H. Leonard, and D.L. Stamp. 1976. Principles of field crop Production. Macmillan, New York, pp 1118.

Martre, P., N. Bertin, C. Salon, and M. Génard. 2011. Modelling the size and composition of fruit, grain and seed by process-based simulation models. New Phytologist 191: 601-618.

Matros, A., S. Kaspar, K. Witzel, and H. P. Mock. 2011. Recent progress in liquid chromatography-based separation and label-free quantitative plant proteomics. Phytochemistry 72: 963-974.

McFadden, E.S., and E.R. Sears. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. Journal of Heredity, 37: 81-89, 107-116.

McIntosh, S., L. Watson, P. Bundock, A. Crawford, J. White, G. Cordeiro, D. Barbary, L. Rooke, and R. Henry. 2007. SAGE of the developing wheat caryopsis. Plant biotechnology journal 5: 69-83.

McLafferty, F. 1981. Tandem mass spectrometry. Science 214: 280-287.

Mhamdi, A., J. Hager, S. Chaouch, G. Queval, Y. Han, L. Taconnat, P. Saindrenan, H. Gouia, E. Issakidis-Bourguet, J.-P. Renou, and G. Noctor. 2010. Arabidopsis GLUTATHIONE REDUCTASE1 Plays a Crucial Role in Leaf Responses to Intracellular Hydrogen Peroxide and in Ensuring Appropriate Gene Expression through Both Salicylic Acid and Jasmonic Acid Signaling Pathways. Plant physiology 153: 1144-1160.

Monti, M., S. Orrù, D. Pagnozzi, and P. Pucci. 2005. Functional proteomics. Clinica Chimica Acta 357: 140-150.

Morrison, I.N. 1976. The structure of the chlorophyll-containing cross cells and tube cells of the inner pericarp of wheat during grain development. Botanical Gazette, The Uni. of Chicago press, 137, pp 85-93.

Müller, S., and H. Wieser. 1995. The location of disulphide bonds in α-type gliadins. Journal of Cereal Science 22: 21-27.

Nadaud, I., C. Girousse, C. Debiton, C. Chambon, M. F. Bouzidi, P. Martre, and G. Branlard. 2010. Proteomic and morphological analysis of early stages of wheat grain development. Proteomics 10: 2901-2910.

Nieto-Taladriz, M. T., M. R. Perretant, and M. Rousset. 1994. Effect of gliadins and HMW and LMW subunits of glutenin on dough properties in the F_6 recombinant inbred lines from a bread wheat cross. TAG Theoretical and Applied Genetics 88: 81-88.

Nonogaki, H., G.W. Bassel, and J.D. Bewley. 2010. Germination—Still a mystery. *Plant Science*. 179:574-581.

Ogihara, Y., K. Mochida, Y. Nemoto, K. Murai, Y. Yamazaki, I. T. Shin, and Y. Kohara. 2003. Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. The Plant journal : for cell and molecular biology 33: 1001-1011.

Olsen, O. A. 2001. Endosperm Development: Cellularization and Cell Fate Specification. Annual review of plant physiology and plant molecular biology 52: 233-267.

Olsen, O. A. 2004. Nuclear endosperm development in cereals and Arabidopsis thaliana. The Plant cell 16 Suppl: S214-227.

Olsen, O. A., C. Linnestad, and S. E. Nichols. 1999. Developmental biology of the cereal endosperm. Trends in plant science 4: 253-257.

Opanowicz, M., P. Hands, D. Betts, M. L. Parker, G. A. Toole, E. N. C. Mills, J. H. Doonan, and S. Drea. 2011. Endosperm development in *Brachypodium distachyon*. Journal of Experimental Botany 62: 735-748.

Osborne, T.B., 1907. The proteins of the wheat kernel. Carnegie Inst: Washington DC Publ, 84: 1-119.

Oury FX, P. Bérard, M. Brancourt-Hulmel, C. Depatureaux, G. Doussinault, N. Galic, A. Giraud, E. Heumez, C. Lecomte, P. Pluchard, B. Rolland, M. Rousset, M. Trottet. **2003**. Yield and grainprotein concentration in bread wheat: a review and a study of multi-annual data from a French breeding program. Journal Genetics and Breeding 57: 59-68

Özkan, H., G. Willcox, A. Graner, F. Salamini, and B. Kilian. 2011. Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). Genetic Resources and Crop Evolution 58: 11-53.

Palmblad, M., D. J. Mills, and L. V. Bindschedler. 2008. Heat-Shock Response in Arabidopsis thaliana Explored by Multiplexed Quantitative Proteomics Using Differential Metabolic Labeling. Journal of Proteome Research 7: 780-785.

Panchaud, A., J. Hansson, M. Affolter, R. Bel Rhlid, S. Piu, P. Moreillon, and M. Kussmann. 2008. ANIBAL, Stable Isotope-based Quantitative Proteomics by Aniline and Benzoic Acid Labeling of Amino and Carboxylic Groups. Molecular and Cellular Proteomics 7: 800-812.

Parisy, V., B. Poinssot, L. Owsianowski, A. Buchala, J. Glazebrook, and F. Mauch. 2007. Identification of PAD2 as a γ-glutamylcysteine synthetase highlights the importance of glutathione in disease resistance of Arabidopsis. The Plant Journal 49: 159-172.

Parker, M. L. 1980. Protein Body Inclusions in Developing Wheat Endosperm. Annals of Botany 46: 29-36.

Parker, M. L. 1982. Protein accumulation in developing endosperm of a high-protein line of Triticum dicoccoides. Plant, Cell & Environment 5: 37-43.

Parker, M. L., A. Ng, and K. W. Waldron. 2005. The phenolic acid and polysaccharide composition of cell walls of bran layers of mature wheat (Triticum aestivum L. cv. Avalon) grains. Journal of the Science of Food and Agriculture 85: 2539-2547.

Paux, E., P. Sourdille, I. Mackay, and C. Feuillet. 2011. Sequence-based marker development in wheat: Advances and applications to breeding. Biotechnology advances.

Payne, P. I. 1987. Genetics of Wheat Storage Proteins and the Effect of Allelic Variation on Bread-Making Quality. Annual Review of Plant Physiology 38: 141-153.

Peltier, J. B., Y. Cai, Q. Sun, V. Zabrouskov, L. Giacomelli, A. Rudella, A. J. Ytterberg, H. Rutschow, and K. J. van Wijk. 2006. The oligomeric stromal proteome of Arabidopsis thaliana chloroplasts. Molecular & cellular proteomics : MCP 5: 114-133.

Percival, J. 1921. The Wheat Plant. Duckworth & Co, London, pp 463.

Peterson, R.F. 1965. Wheat: Botany, cultivation and utilisation. Leonard Hill, London, pp 448.

Philippe, R., F. Choulet, E. Paux, J. van Oeveren, J. Tang, A. H. Wittenberg, A. Janssen, M. J. van Eijk, K. Stormo, A. Alberti, P. Wincker, E. Akhunov, E. van der Vossen, and C. Feuillet. 2012. Whole Genome Profiling provides a robust framework for physical mapping and sequencing in the highly complex and repetitive wheat genome. BMC genomics 13: 47.

Pichler, P., T. Köcher, J. Holzmann, T. Möhring, G. Ammerer, and K. Mechtler. 2011. Improved Precision of iTRAQ and TMT Quantification by an Axial Extraction Field in an Orbitrap HCD Cell. Analytical Chemistry 83: 1469-1474.

Pomeranz, Y. 1988. Chemical composition of kernel structures. In: Pomeranz Y (ed) Wheat: chemistry and technology, 3rd edition, AACC. St Paul MN, USA, pp 97-158.

Popineau, Y., and Pineau, F. 1988. Changes of conformation and surface hydrophobicity of gliadins. Vol. 21. Kidlington, ROYAUME-UNI: Elsevier.

Posner, E.S. 2000. Wheat. In: Kulp, K., *et al.* (Eds.) Handbook of cereal science and technology, cop. New York: Marcel Dekker, pp 1-99.

Rabilloud, T. 2002. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2: 3-10.

Rajjou, L., M. Duval, K. Gallardo, J. Catusse, J. Bally, C. Job, and D. Job. 2012. Seed germination and vigor. *Annu Rev Plant Biol*. 63:507-533.

Romeuf, I., 2010. Identification in silico des facteurs de transcription du blé tendre (*Triticum aestivum*) et mise en évidence des facteurs de transcription impliqués dans la synthèse des protéines de réserve. Ph.D. thesis, Université Blaise Pascal, Clermont-Ferrand, France. pp205.

Ross, P. L. 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3: 1154-1169.

Rost, T.L., and N.R. Lersten. 1973. A synopsis and selected bibliography of grass caryopsis anatomy and fine structure. Iowa State, J Res 48: 47-87.

Sabelli, P. A., and B. A. Larkins. 2009. The Development of Endosperm in Grasses. Plant physiology 149: 14-26.

Salse, J., S. Bolot, M. Throude, V. Jouffe, B. Piegu, U. M. Quraishi, T. Calcagno, R. Cooke, M. Delseny, and C. Feuillet. 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. The Plant cell 20: 11-24.

Salse, J., M. Abrouk, S. Bolot, N. Guilhot, E. Courcelle, T. Faraut, R. Waugh, T. J. Close, J. Messing, and C. Feuillet. 2009. Reconstruction of monocotelydoneous proto-chromosomes reveals faster evolution in plants than in animals. Proceedings of the National Academy of Sciences of the United States of America 106: 14908-14913.

Saulnier, L., F. Guillon, and A.-L. Chateigner-Boutin. 2012. Cell wall deposition and metabolism in wheat grain. Journal of Cereal Science.

Sax, K., 1922. Sterility in wheat hybrids. II. Chromosome behavior in partially sterile hybrids. Genetics, 7: 513-550.

Schaff, J. E., F. Mbeunkui, K. Blackburn, D. M. Bird, and M. B. Goshe. 2008. SILIP: a novel stable isotope labeling method for in planta quantitative proteomic analysis. The Plant Journal 56: 840-854.

Schlaeppi, K., N. Bodenhausen, A. Buchala, F. Mauch, and P. Reymond. 2008. The glutathione-deficient mutant pad2-1 accumulates lower amounts of glucosinolates and is more susceptible to the insect herbivore Spodoptera littoralis. The Plant Journal 55: 774-786.

Schnyder, and U. Baum. 1992 Growth of the grain of wheat (*Triticum aestivum* L.). The relationship between water content and dry matter accumulation. Eur J Agron 1: 51-57.

Schütz, W., N. Hausmann, K. Krug, R. Hampp, and B. Macek. 2011. Extending SILAC to Proteomics of Plant Cell Lines. The Plant Cell Online 23: 1701-1705.

Schwartz, E. 2007. Characterization of growing microorganisms in soil by stable isotope probing with H218O. Applied and environmental microbiology 73: 2541-2546.

She, M., X. Ye, Y. Yan, C. Howit, M. Belgard, and W. Ma. 2011. Gene networks in the synthesis and deposition of protein polymers during grain development of wheat. Functional & Integrative Genomics 11: 23-35.

Shewry, P. R., A. S. Tatham, J. Forde, M. Kreis, and B. J. Miflin. 1986. The classification and nomenclature of wheat gluten proteins: A reassessment. Journal of Cereal Science 4: 97-106.

Shewry, P. R., and A. S. Tatham. 1997. Disulphide Bonds in Wheat Gluten Proteins. Journal of Cereal Science 25: 207-227.

Shewry, P.R., A.S. Tatham, 1999. The characteristics, structures and evolutionary relationships of prolamins. In: Seed Proteins, pp. 11–33. Eds. Shewry, P.R. and Casey, R. Kluwer Academic Publishers, Dorfrecht, the Netherlands.

Shewry, P.R., Halford, N.G. 2003. Genetics of wheat gluten proteins. In Advances in Genetics, Vol 49, Vol 49. Academic Press, San Diego, pp 111-184.

Shewry, P. R. 2009. Wheat. Journal of Experimental Botany 60: 1537-1553.

Simmonds, D. H. and T. P. O'Brien. **1981**. Morphological and biochemical development of the wheat endosperm. In Advances in Cereal Science and Technology; Pomeranz, Y., Ed.; AACC: St. Paul, MN; vol 4, pp 5−70.

Singh, J., and J. H. Skerritt. 2001. Chromosomal Control of Albumins and Globulins in Wheat Grain Assessed using Different Fractionation Procedures. Journal of Cereal Science 33: 163-181.

Soltner, D. 1988. Les grandes productions végétales: céréales, plantes sarclées, prairies: Collection Sciences et Techniques Agricoles.

Spannagl, M., K. Mayer, J. Durner, G. Haberer, and A. Frohlich. 2011. Exploring the genomes: from Arabidopsis to crops. Journal of plant physiology 168: 3-8.

Šramková, Z., E. Gregová, and E. Šturdík. 2009. Chemical composition and nutritional quality of wheat grain. Acta Chimica Slovaca, 2, 115–138.

Sreenivasulu, N., V. Radchuk, M. Strickert, O. Miersch, W. Weschke, and U. Wobus. 2006. Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death

and ABA- regulated maturation in developing barley seeds. The Plant journal : for cell and molecular biology 47: 310-327.

Surget, A., and C. Barron. 2005. Histologie du grain de blé. Industries des Céréales, 145: 3-7.

Szabó, A.T., K. Hammer. 1996. Notes on the taxonomy of farro: *Triticum monococcum, T. dicoccon* and *T. spelta. Hulled wheats.* Proceedings of the 1st International Workshop on Hulled Wheats, 21–22 July 1995, Tuscany, Italy. IPGRI, Italy, pp 2–40.

Takemoto, Y., S. J. Coughlan, T. W. Okita, H. Satoh, M. Ogawa, and T. Kumamaru. 2002. The rice mutant esp2 greatly accumulates the glutelin precursor and deletes the protein disulfide isomerase. Plant physiology 128: 1212-1222.

Tian, L., G. Peel, Z. Lei, N. Aziz, X. Dai, J. He, B. Watson, P. Zhao, L. Sumner, and R. Dixon. 2009. Transcript and proteomic analysis of developing white lupin (*Lupinus albus* L.) roots. BMC plant biology 9.

Tosi, P., M. Parker, C. S. Gritsch, R. Carzaniga, B. Martin, and P. R. Shewry. 2009. Trafficking of storage proteins in developing grain of wheat. Journal of Experimental Botany 60: 979-991.

Tranquilli, G., and J. Dubcovsky. 2000. Epistatic interaction between vernalization genes Vrn-Am1 and Vrn-Am2 in diploid wheat. Journal of Heredity 91: 304-306.

Ünlü, M., M. E. Morgan, and J. S. Minden. 1997. Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. Electrophoresis 18: 2071-2077.

Vavilov, N.I. 1940. The theory of the origin of cultivated plants after Darwin. Soviet Science 2: 55-75.

Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial Analysis of Gene Expression. Science 270: 484-487.

Vensel, W. H., C. K. Tanaka, N. Cai, J. H. Wong, B. B. Buchanan, and W. J. Hurkman. 2005. Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics 5: 1594-1611.

Volkmuth, W., S. Turk, A. Shapiro, Y. Fang, E. Kiegle, M. van Haaren, and J. Donson. 2003. Technical advances: genome-wide cDNA-AFLP analysis of the Arabidopsis transcriptome. Omics : a journal of integrative biology 7: 143-159.

Wan, Y., R. L. Poole, A. K. Huttly, C. Toscano-Underwood, K. Feeney, S. Welham, M. J. Gooding, C. Mills, K. J. Edwards, P. R. Shewry, and R. A. Mitchell. 2008. Transcriptome analysis of grain development in hexaploid wheat. BMC genomics 9: 121.

Wang, K., X. L. An, L. P. Pan, K. Dong, L. Y. Gao, S. L. Wang, Z. Z. Xie, Z. Zhang, R. Appels, W. Ma, and Y. M. Yan. 2012. Molecular characterization of HMW-GS 1Dx3t and 1Dx4t genes from Aegilops tauschii and their potential value for wheat quality improvement. Hereditas 149: 41-49.

Ward, J. A., L. Ponnala, and C. A. Weber. 2012. Strategies for transcriptome analysis in nonmodel plants. American Journal of Botany 99: 267-276.

Waters, K. M., J. G. Pounds, and B. D. Thrall. 2006. Data merging for integrated microarray and proteomic analysis. Briefings in Functional Genomics & Proteomics 5: 261-272.

Whitelaw, C. A., W. B. Barbazuk, G. Pertea, A. P. Chan, F. Cheung, Y. Lee, L. Zheng, S. van Heeringen, S. Karamycheva, J. L. Bennetzen, P. SanMiguel, N. Lakey, J. Bedell, Y. Yuan, M. A. Budiman, A. Resnick, S. Van Aken, T. Utterback, S. Riedmuller, M. Williams, T. Feldblyum, K. Schubert, R. Beachy, C. M. Fraser, and J. Quackenbush. 2003. Enrichment of gene-coding sequences in maize by genome filtration. Science 302: 2118-2120.

Whitelegge, J. P., J. E. Katz, K. A. Pihakari, R. Hale, R. Aguilera, S. M. Gomez, K. F. Faull, D. Vavilin, and W. Vermaas. 2004. Subtle modification of isotope ratio proteomics; an integrated strategy for expression proteomics. Phytochemistry 65: 1507-1515.

Wilkins, M. R., C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J.-C. Sanchez, J. X. Yan, A. A. Gooley, G. Hughes, I. Humphery-Smith, K. L. Williams, and D. F. Hochstrasser. 1996. From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. Nat Biotech 14: 61-65.

Wilson, I. D., G. L. Barker, R. W. Beswick, S. K. Shepherd, C. Lu, J. A. Coghill, D. Edwards, P. Owen, R. Lyons, J. S. Parker, J. R. Lenton, M. J. Holdsworth, P. R. Shewry, and K. J.

Edwards. 2004. A transcriptomics resource for wheat functional genomics. Plant biotechnology journal 2: 495-506.

Winfield, M. O., C. Lu, I. D. Wilson, J. A. Coghill, and K. J. Edwards. 2009. Cold- and light-induced changes in the transcriptome of wheat leading to phase transition from vegetative to reproductive growth. BMC plant biology 9: 55.

Winton, A.L., and Winton, Kate B. 1932. Structure and composition of foods. By Vol. I. Cereals, Starch, Oil Seeds, Nuts, Oils, Forage Plants. Pp. xiv +710. New York: J. Wiley & Sons; London: Chapman & Hall. 53s. Journal of the Society of Chemical Industry 51: 539-540.

Worland, T., and J.W. Snape. 2001. Genetic basis of worldwide wheat varietal improvement. In: Bonjean, A.P. and Angus, W.J. (Eds.), pp. 84- 85. The world wheat book: A history of wheat breeding. Lavoisier Publishing Inc., France.

Xu, S. B., H. T. Yu, L. F. Yan, and T. Wang. 2010. Integrated Proteomic and Cytological Study of Rice Endosperms at the Storage Phase. Journal of Proteome Research 9: 4906-4918.

Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, and J. Dubcovsky. 2003. Positional cloning of the wheat vernalization gene VRN1. Proceedings of the National Academy of Sciences of the United States of America 100: 6263-6268.

Yan, L., A. Loukoianov, A. Blechl, G. Tranquilli, W. Ramakrishna, P. SanMiguel, J. L. Bennetzen, V. Echenique, and J. Dubcovsky. 2004. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303: 1640-1644.

Yan, L., D. Fu, C. Li, A. Blechl, G. Tranquilli, M. Bonafede, A. Sanchez, M. Valarik, S. Yasuda, and J. Dubcovsky. 2006. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proceedings of the National Academy of Sciences of the United States of America 103: 19581-19586.

Young, T. E., and D. R. Gallie. 2000. Programmed cell death during endosperm development. Plant Molecular Biology 44: 283-301.

Zadoks, J. C., T. T. Chang, and C. F. Konzak. 1974. A decimal code for the growth stages of cereals. Weed Research 14: 415-421.

Zhang, W., F. Li, and L. Nie. 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. Microbiology 156: 287-301.

Zhao, S., and W. B. Bruce. 2003. Expression profiling using cDNA microarrays. Methods in molecular biology (Clifton, N.J.) 236: 365-380.

ANNEXES

ANNEX 1: Supplemental Information data of Article 1, Proteomic analysis of peripheral layers during wheat (*Triticum aestivum* L.) grain development

SupplInfo Table 1: Proteins of wheat grain PLs development identified by MALDI-TOF or MS/MS and classified by their corresponding metabolic pathway and biochemical functions. Spots ID with asterix (*) represent proteins of redox homeostasis

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Coverage rate	Number of matched peptides	Experimental pI/MW (Da)	Theoratical pI/MW (Da)	MS/MS peptides sequence with valid score
PROFILE 2												
METABOLISM												
Carbohydrate metabolism												
1193	gi|1174745		viridiplantae	Secale cereale	Triosephosphate isomerase	74	0.037	0.42	9	6/31613	5.16/30679	
456	NP_001105211	CK207916	wh	Zea mays	Fructokinase 2	90	5.80e-03	0.52	12	5.34/35482	5.01/43973	
140	gi|19106		viridiplantae	Hordeum vulgare subsp. vulgare	Sucrose synthase	164	3.80e-11	0.31	23	5.94/92153	5.5/72991	
934	gi|32478662		viridiplantae	Triticum aestivum	Cytosolic Glyceraldehyde-3-phosphate dehydrogenase	93	4.40e-04	0.66	10	6.34/18167	7.03/44923	
667	gi|1174745		viridiplantae	Secale cereale	Triosephosphate isomerase	99	1.20e-04	0.48	11	6/31613	5.07/30565	
1198	gi|148508784		viridiplantae	Triticum aestivum	Glyceraldehyde-3-phosphate dehydrogenase	161	7.7e-11	0.58	17	7.08/36626	6.64/41780	
Energy metabolism												
334*	gi|20302473		viridiplantae	Triticum aestivum	Ferredoxin-NADP(H) oxidoreductase	123	4.80e-07	0.45	17	6.92/40206	5.13/42335	
733	gi|54303886		viridiplantae	Molinia caerulea	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	77	0.018	0.24	12	6.13/46678	5.1/48017	
120	gi|31087891		viridiplantae	Hordeum patagonicum subsp. santacrucense	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	91	8.20e-04	0.24	15	5.95/52739	4.96/45597	
12	gi|134102		viridiplantae	Triticum aestivum	RuBisCO	195	3.10e-14	0.55	21	4.83/57485	4.84/62371	
Amino acid metabolism												
817	gi|68655495		viridiplantae	Hordeum vulgare subsp. vulgare	Methionine synthase 1 enzyme	100	1.00e-04	0.29	15	5.74/84511	5.54/68377	
Metabolism of other amino acid												
1410*	gi|23504745		viridiplantae	Triticum aestivum	Glutathione transferase F5	104	3.80e-05	0.44	7	5.78/23422	5.54/27431	
465	gi|40317416		viridiplantae	Triticum aestivum	Glutamine synthetase isoform GSr1	80	9.40e-03	0.23	7	5.35/38705	5.19/49235	
Biosynthesis of secondary metabolites												
697	BAC98588	CD931410	wh	Oryza sativa Japonica Group	Putative 2-oxoGlutarate-dependent dioxygenase	85	0.019	0.4	6	5.36/33983	4.97/40585	
GENETIC INFORMATION AND PROCESSING												
Protein synthesis												
1136	AAL79732	gi|22652	viridiplantae	Oryza sativa	Heat shock protein 90	144	3.80e-09	0.28	21	4.89/93045	4.87/77315	
165	gi|2827002		viridiplantae	Triticum aestivum	HSP70	117	1.90e-06	0.26	12	5.14/70986	4.67/68769	
150	gi|476003		viridiplantae	Hordeum vulgare subsp. vulgare	HSP70	107	1.90e-05	0.24	11	5.76/66975	4.54/67088	
444	XP_002526446	gi|148907083	viridiplantae	Triticum aestivum	Heat shock protein, putative	76	0.027	0.14	8	5.35/75363	4.73/71207	
391	gi|556673		viridiplantae	Secale cereale	Heat-shock protein	73	0.044	0.11	6	4.9/88063	4.73/74703	
Protein Folding												
288	gi|12056115		viridiplantae	Triticum turgidum subsp. durum	Protein disulfide isomerase	82	5.70e-03	0.18	8	4.99/56544	4.94/61503	
521	BAD35228	CJ594539	wh	Oryza sativa Japonica Group	Putative chaperonin 21 precursor	82	0.036	0.41	6	7.72/26319	5.12/29477	
Protein Turn over												
1167	gi|195641684		viridiplantae	Zea mays	Proteasome subunit alpha type 6	94	4.30e-04	0.32	8	6.11/27470	6.1/30616	
Translation												
1363	NP_001147501	CJ730434	wh	Zea mays	Eukaryotic translation initiation factor 3	96	1.70e-03	0.41	9	6/36260	6.44/42836	
1042	ACA50518	CJ798708	wh	Oryza sativa Japonica Group	Elongation factor 1-gamma	117	1.20e-05	0.51	10	6.31/46914	6.38/49763	
613	ACA50518	CJ630041	wh	Oryza sativa Japonica Group	Elongation factor 1-gamma	257	1.80e-07	0.18	3	6.31/46914	5.96/52071	YQEENTVSYVTLNK+2, LGY**M**PYNAQSEEFGIAGlK +2 Oxidation (M)
679	AAF15312	CJ721551	wh	Oryza sativa	Chloroplast translational elongation factor Tu	82	0.037	0.43	7	6.05/50355	5.18/52920	

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Coverage rate	Number of matched peptides	Experimental pI/MW (Da)	Theoratical pI/MW (Da)	MS/MS peptides sequence with valid score
Transcription												
96	ACG34122	CJ731020	wh	Zea mays	Ribonucleoprotein	101	5.00e-04	0.53	8	4.44/31790	4.44/31509	
176	ACG34122	CJ731020	wh	Zea mays	Ribonucleoprotein	98	0.001	0.43	7	4.44/31790	4.48/32275	
ENVIRONMENTAL INFORMATION AND PROCESSING												
Signal transduction												
887	gi|195622544		viridiplantae	Zea mays	14-3-3-like protein	78	0.016	0.35	9	4.75/29480	4.77/34969	
ATP INTERCONVERSION												
924	AAF91407	CJ787049	wh	Lolium perenne	Nucleoside diphosphate kinase	113	3.10e-05	0.47	10	6.3/16501	6.02/16636	
1320	gi|194033257		viridiplantae	Oryza sativa Japonica Group	ATP synthase F0 subunit 1	73	0.053	0.25	11	5.85/55339	5.36/60037	
STRESS DEFENSE												
437*	gi|15808779		viridiplantae	Hordeum vulgare subsp. vulgare	Ascorbate peroxidase	81	8.20e-03	0.37	7	5.1/27622	4.97/31377	
UNKNOWN / HYPOTHETICAL PROTEIN												
86	EAZ05175	EG376537	poa	Oryza sativa Indica Group	Hypothetical protein	147	4.20e-08	0.67	16	7.29/39170	6.46/39772	
910	gi|219363507		viridiplantae	Zea mays	Hypothetical protein	75	0.032	0.17	6	6.58/46069	5.85/50320	
414	CAA61258	CV774605	wh	Zea mays	Orf	89	7.70e-03	0.41	10	5.29/45746	4.84/52839	
1211	EAZ41123	CD909834	wh	Oryza sativa Japonica Group	Hypothetical protein	90	5.60e-03	0.74	10	7.29/39200	6.6/41605	
PROFILE 3												
METABOLISM												
Carbohydrate metabolism												
725	BAA22419	gi|218196777	viridiplantae	Oryza sativa Japonica Group	Orthophosphate dikinase	80	0.012	0.17	18	5.98/102802	5.13/77096	
1315	CAZ76054	gi|120680	viridiplantae	Triticum aestivum	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	116	2.80e-06	0.42	15	6.67/36528	6.37/45644	
1504	CAC85913	gi|18076790	viridiplantae	Triticum aestivum	PhosphoGlucomutase	164	4.50e-11	0.41	19	5.66/62750	5.58/64999	
1401	ABB46862	gi|222624569	viridiplantae	Oryza sativa Japonica Group	Enolase, putative, expressed	74	0.044	0.29	11	5.72/51583	5.19/58233	
1464	CAA74160	CJ787667	wh	Hordeum vulgare	Alpha-galactosidase	92	3.70e-03	0.44	11	6.44/22053	6.18/45695	
Storage protein												
1187	ABF95817	BJ288636	wh	Oryza sativa Japonica Group	Cupin family protein, expressed	87	0.011	0.6	11	7.18/61439	8.82/36232	
976	ACJ65514	gi|215398471	viridiplantae	Triticum aestivum	Globulin 3	120	3.0e-04	0.05	4	7.78/66310	4.84/19253	GSSNLQVV**C**FEINAER +2Carbamidomethyl (C), GSSNLQVV**C**FEINAER +2 Carbamidomethyl (C)
Energy metabolism												
796	gi|57283816		viridiplantae	Dactylis Glomerata	Ribulose bisphosphate carboxylase large chain	121	8.90e-07	0.28	17	6.04/51203	6.17/35304	
1172*	gi|21263612		viridiplantae	Hordeum vulgare	NAD-dependent formate dehydrogenase	140	1.10e-08	0.42	17	6.90/41519	6.13/49524	
Lipid metabolism												
1446	BAE20413	CJ683541	wh	Triticum aestivum	Aspartic proteinase	81	0.047	0.49	11	5.14/54315	4.78/31503	
1477	gi|73912434		viridiplantae	Triticum aestivum	Aspartic proteinase	167	5.90e-07	0.06	2	5.14/54280	4.74/17489	KFALKPEEYILK +2, VGEGAAAQ**C**ISGFTA**M**DIPPPR +2 Carbamidomethyl (C),Oxidation (M)
Amino acid metabolism												
822	P52894	CJ780425	wh	Hordeum vulgare	Alanine aminotransferase 2	86	0.016	0.45	8	5.93/52877	5.98/55395	
GENETIC INFORMATION AND PROCESSING												
Protein synthesis												
1441	gi|2827002		viridiplantae	Triticum aestivum	HSP70	86	2.60e-03	0.29	15	5.14/70986	5.96/50029	
1502	gi|2827002		viridiplantae	Triticum aestivum	HSP70	123	5.60e-07	0.3	20	5.14/70986	5.02/68232	
899	ABA95501	CJ778807	wh	Oryza sativa (indica group)	Heat shock cognate 70 kDa protein, putative	81	0.047	0.37	7	4.97/67320	4.95/68554	

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Coverage rate	Number of matched peptides	Experimental pI/MW (Da)	Theoratical pI/MW (Da)	MS/MS peptides sequence with valid score
ENVIRONMENTAL INFORMATION AND PROCESSING												
Membrane transport												
1364	ACG48482	BJ291619	wh	Zea mays	Lipoprotein	80	0.063	0.42	7	5.87/26214	4.78/48230	
STRESS DEFENSE												
1321*	BAB71741	CJ671521	wh	Oryza sativa Japonica Group	Glyoxalase I	114	2.50e-05	0.5	16	5.15/32539	5.32/39607	
1500*	BAB71741	CN013057	wh	Oryza sativa Japonica Group	Glyoxalase I	82	0.043	0.48	11	5.15/32539	5.21/40223	
UNKNOWN / HYPOTHETICAL PROTEIN												
1519	EEE55430	BJ232559	wh	Oryza sativa Japonica Group	Hypothetical protein OsJ_03566	113	0.016	0.13	2	5.2/49726	5.25/56507	SLDFAAPHYLSGPLR +2, EHGGGFLTDHFPSAR +2
PROFILE 4												
METABOLISM												
Carbohydrate metabolism												
689	CAZ76054		viridiplantae	Triticum aestivum	Glyceraldehyde- 3-Phosphate Dehydrogenase	72	0.061	0.18	8	6.67/36528	5.37/50336	
460	gi|8918504		viridiplantae	Triticum aestivum	Glucose-6-phosphate dehydrogenase	110	9.70e-06	0.19	11	5.92/58286	5.86/60828	
513	gi|8272480		viridiplantae	Avena sativa	Fructose 1,6-bisphosphate aldolase precursor	76	0.023	0.17	6	9.01/41895	5.13/43831	
992	gi|37928995		viridiplantae	Triticum aestivum	Cytosolic malate dehydrogenase	143	4.80e-09	0.73	15	6.6/24332	5.54/44332	
147	gi|2507469		viridiplantae	Hordeum vulgare	Triosephosphate isomerase	82	6.20e-03	0.33	7	5.39/26720	4.81/62676	
186	gi|8272480		viridiplantae	Avena sativa	Fructose 1,6-bisphosphate aldolase precursor	85	3.10e-03	0.22	6	9.01/41895	5.07/42863	
328	ACM78035	DV478664	poa	Triticum aestivum	Chloroplast fructose-bisphosphate aldolase	88	0.035	0.37	8	5.94/42015	5.2/43215	
Energy metabolism												
184	AAO33154	BJ267613	wh	Oryza sativa Japonica Group	Putative transketolase	81	0.046	0.44	5	6.12/80029	5.24/69208	
333	AAO33154	BJ267613	wh	Oryza sativa Japonica Group	Putative transketolase	97	0.0013	0.58	9	6.12/80029	5.29/69057	
384	AAO33154	CJ719370	wh	Oryza sativa Japonica Group	Putative transketolase	80	0.063	0.38	7	6.12/80029	4.94/53858	
540	gi|4038721		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	99	1.20e-04	0.57	10	8.83/18533	5.25/17511	
581	gi|544700		viridiplantae	Hordeum vulgare	Light-harvesting complex I	204	1.30e-04	0.1	3	8.13/24211	4.93/26537	YPGGAFDPLGFSK +2, FKESEIYHCR +2 Carbamidomethyl (C), KYPGGAFDPLGFSK +2
612	ACG28457	BE591357	wh	Zea mays	Chlorophyll a-b binding protein 8	80	0.06	0.46	6	8.94/28984	5.07/29271	
448	Q7X9A6	CJ897844	wh	Triticum aestivum	Cytochrome b6-f complex iron-sulfur subunit.	91	0.0045	0.53	10	8.47/23726	5.62/22122	
455	gi|61378618		viridiplantae	Triticum monococcum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	152	6.10e-10	0.36	20	6.04/52716	4.52/55834	
23	gi|134290407		viridiplantae	Triticum aestivum	Putative oxygen-evolving complex precursor	78	0.014	0.41	9	9.72/21126	8.58/18223	
26	gi|14017580		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	204	3.80e-15	0.44	25	6.22/52817	4.63/55952	
35	gi|11990897		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	145	3.10e-09	0.66	15	8.8/19448	5.59/17121	
41	gi|18958		viridiplantae	Hordeum vulgare subsp. vulgare	Chlorophyll a/b binding (CAB) protein of photosystem II (PSII)	82	0.007	0.3	7	5.33/30725	4.92/31217	
50	gi|169666638		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	176	2.40e-12	0.39	22	6.22/52817	4.73/56202	
1529	AAB82711	CJ906759	wh	x Tritordeum sp.	Glycine decarboxylase P subunit	88	0.0093	0.33	7	6.32/111020	5.50/71030	
16	gi|13310805		viridiplantae	Eriachne triodioides	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	81	0.0075	0.28	14	6.22/46259	4.55/46086	
18	gi|14017580		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	255	3.10e-20	0.47	27	6.22/52817	4.66/55869	
19	gi|147945622		viridiplantae	Leymus chinensis	Chloroplast oxygen-evolving enhancer protein 1	112	6.10e-06	0.44	13	6.08/34490	5/35480	
21	gi|119710748		viridiplantae	Thysananthus anguiformis	Ribulose 1,5-bisphosphate carboxylase/oxygenase	75	0.031	0.3	7	8.86/34295	4.96/46926	
103	NP_001148439	CJ930673	wh	Zea mays	Chlorophyll a-b binding protein 2	102	3.90e-04	0.33	7	5.46/27945	4.88/30413	
141	gi|14017580		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	106	2.40e-05	0.31	17	6.22/52817	4.69/46378	
153	gi|11990897		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	113	4.80e-06	0.66	14	8.8/19448	5.73/16156	

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Coverage rate	Number of matched peptides	Experimental pI/MW (Da)	Theoratical pI/MW (Da)	MS/MS peptides sequence with valid score
Metabolism of other amino acid												
1041*	CAD29478	BJ315788	wh	Triticum aestivum	Glutathione transferase F5	84	0.027	0.47	6	5.78/23437	5.92/26716	
344	gi|68655495		viridiplantae	Hordeum vulgare subsp. vulgare	Methionine synthase 1 enzyme	101	7.70e-05	0.2	9	5.74/84511	5.44/70938	
783*	gi|22022400		viridiplantae	Triticum aestivum	Glutathione-S-transferase	83	5.10e-03	0.18	5	5.99/24901	6.05/32019	
GENETIC INFORMATION AND PROCESSING												
Protein folding												
701	BAD35232	CJ694211	wh	Oryza sativa Japonica Group	Putative chaperonin 21 precursor	90	5.60e-03	0.7	9	4.58/21046	6.17/44909	
Translation												
197	NP_001149356	BF482295	wh	Zea mays	Plastid-specific 30S ribosomal protein 2	83	0.033	0.47	8	5.54/16481	4.97/28784	
ENVIRONMENTAL INFORMATION AND PROCESSING												
Signal transduction												
1228	gi|439586		viridiplantae	Hordeum vulgare subsp. vulgare	Calreticulin	80	0.011	0.28	11	4.45/47009	4.61/58010	
ATP Interconversion												
580	gi|14017579		viridiplantae	Triticum aestivum	ATP synthase	150	9.70e-10	0.45	20	5.06/53824	4.97/62937	
629	gi|60391822		viridiplantae	Hordeum vulgare	ATP synthase	76	0.022	0.65	6	5.2/15208	5.19/17809	
982	gi|14017569		viridiplantae	Triticum aestivum	ATP synthase	206	2.40e-15	0.44	22	6.11/55261	6.05/59011	
408	gi|14017579		viridiplantae	Triticum aestivum	ATP synthase	178	1.50e-12	0.49	22	5.06/53824	4.92/58927	
83	gi|14017569		viridiplantae	Triticum aestivum	ATP synthase	187	1.90e-13	0.43	22	6.11/55261	5.8/58066	
STRESS DEFENSE												
7	gi|63021412		viridiplantae	Triticum aestivum	Salt tolerant protein	89	1.10e-03	0.54	11	4.71/17055	4.59/20788	
20	gi|63021412		viridiplantae	Triticum aestivum	Salt tolerant protein	88	1.50e-03	0.49	11	4.71/17055	4.63/19476	
974*	gi|3328221		viridiplantae	Secale cereale	Thioredoxin peroxidase	108	1.50e-05	0.41	7	6.34/28115	4.86/27843	
723*	gi|1805351		viridiplantae	Triticum aestivum	Thiol-specific antioxidant protein	85	2.80e-03	0.54	9	5.71/23312	4.8/27425	
277*	gi|861010		viridiplantae	Hordeum vulgare subsp. vulgare	2-Cys peroxiredoxin	106	2.40e-05	0.43	6	5.48/23284	4.73/27686	
PROFILE 5												
METABOLISM												
Carbohydrate metabolism												
463	gi|7431022		viridiplantae	Triticum aestivum	Glucose and ribitol dehydrogenase	76	0.029	0.33	9	6.54/31627	6/37609	
466	gi|7431022		viridiplantae	Hordeum vulgare	Glucose and ribitol dehydrogenase	80	0.01	0.34	9	6.54/31627	8.84/54450	
516	gi|113595		viridiplantae	Hordeum vulgare	Aldose reductase	73	0.06	0.36	11	6.51/35784	6.59/38913	
585	T06212	BQ805036	wh	Hordeum vulgare	Glucose and ribitol dehydrogenase homolog	120	6.30e-06	0.6	12	6.54/31647	6.61/40119	
1195	CAA74160	CJ787667	wh	Hordeum vulgare	Alpha-galactosidase.	93	0.0029	0.44	12	6.4/22053	6.07/45623	
1252	NP_001105292	BJ294714	wh	Hordeum vulgare	NADP-dependent malic enzyme	85	0.018	0.43	7	7.64/71637	5.9/62947	
459	gi|7431022		viridiplantae	Hordeum vulgare	Glucose and ribitol dehydrogenase homolog	362	2.30e-07	0.25	5	6.54/31627	6.26/39430	ALS**M**QLAEK +2 Oxidation (M), QFGSEVP**M**KR +2 Oxidation (M), GHEDKDAEETLQALR +2, H**M**GPGSSIINTTSVNAYK +2 Oxidation (M), VNGVAPGPIWTPLIPASFPEEK +2
1291	gi|19851522		viridiplantae	Zea mays	Pyruvate decarboxylase	147	2.10e-06	0.06	2	5.72/65021	5.58/62554	ILHHTIGIPDFSQELR +2, MLTGDSAVIAETGDSWFN**C**QK +2 Carbamidomethyl (C)
626	gi|108708038		viridiplantae	Oryza sativa Japonica Group	Fumarate hydratase 1, mitochondrial precursor, putative, expressed	101	8.90e-05	0.26	14	6.93/53682	6.47/52719	
Storage proteins												
111	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	202	7.10e-15	0.41	23	7.78/66310	8.33/55908	
1164	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	80	0.011	0.15	9	7.78/66310	5.06/20145	

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Coverage rate	Number of matched peptides	Experimental pI/MW (Da)	Theoratical pI/MW (Da)	MS/MS peptides sequence with valid score
1352	ACJ65514	CD910651	wh	Triticum aestivum	Globulin 3	145	2.00e-08	0.63	13	7.78/66350	5.59/42646	
137	1802402A	BQ842018	wh	Zea mays	Globulin 2	152	3.90e-09	0.65	15	6.16/49923	6.72/57462	
1387	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	75	0.038	0.24	12	7.78/66310	5.72/36679	
145	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	375	9.90e-07	0.15	5	7.78/66310	5.1/20435	RGSGSESEEEQDQQR +2, GSSNLQVVCFEINAER +2 Carbamidomethyl (C), GSAFVVPPGHPVVEIASSR +2, DQQDEGFVAGPEQQQEHER +2, AKDQQDEGFVAGPEQQQEHER +2
1512	ACJ65515	CD908963	wh	Triticum aestivum	Globulin 3B	123	3.10e-06	0.55	12	7.36/56931	5.92/41810	
218	ACJ65513	BQ579030	wh	Triticum aestivum	Globulin 3c	107	1.20e-04	0.45	12	9.15/38429	6.24/18074	
229	ACJ65514	BQ579386	wh	Triticum aestivum	Globulin 3.	109	7.90e-05	0.59	11	7.78/66350	8.4/45766	
241	ACJ65514	CD917192	wh	Triticum aestivum	Globulin 3	107	1.20e-04	0.56	11	7.78/66350	5.84/18491	
248	1802402A	BQ842018	wh	Zea mays	Globulin 2	151	5.00e-09	0.65	14	6.16/49923	6.56/58518	
255	1802402A	AL818836	wh	Zea mays	Globulin 2	94	2.20e-03	0.42	6	6.16/49923	6.28/58096	
270	ACJ65513	BQ579030	wh	Triticum aestivum	Globulin 3C	111	5.00e-05	0.45	12	9.15/38429	6.66/57857	
278	AAM33459	BQ246191	wh	Oryza sativa Japonica Group	Globulin like protein	115	2.00e-05	0.51	13	6.79/52066	6.24/18484	
320	AAM33459	BQ842018	wh	Oryza sativa Japonica Group	Globulin like protein	138	9.90e-08	0.63	14	6.79/52066	6.68/58843	
321	AAM33459	BQ246191	wh	Oryza sativa Japonica Group	Globulin like protein	146	1.60e-08	0.59	15	6.79/52066	6.39/57494	
33	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	91	8.70e-04	0.25	15	7.78/66310	7.23/57561	
36	gi|215398472		viridiplantae	Triticum aestivum	Globulin 3B	105	3.50e-05	0.24	16	7.36/56897	6.93/57166	
374	ACJ65514	CD910493	wh	Triticum aestivum	Globulin 3	127	1.20e-06	0.51	13	7.78/66350	5.48/42195	
385	ACJ65514	CD906135	wh	Triticum aestivum	Globulin 3	101	5.00e-04	0.62	10	7.78/66350	4.86/19889	
409	ACJ65514	CD907226	wh	Triticum aestivum	Globulin 3	94	2.30e-03	0.5	8	7.78/66350	4.97/21526	
411	ACJ65514	BI479035	wh	Triticum aestivum	Globulin 3	101	5.00e-04	0.71	9	7.78/66350	4.81/19889	
44	ACJ65514	CD910470	wh	Triticum aestivum	Globulin 3	119	7.90e-06	0.6	12	7.78/66350	5.74/43714	
469	ACJ65514	BQ806157	wh	Triticum aestivum	Globulin 3	117	1.20e-05	0.54	11	7.78/66350	5.85/42105	
49	AAM33459	CD918433	wh	Oryza sativa Japonica Group	Globulin like protein	110	6.30e-05	0.49	12	6.79/52066	6.52/57494	
1536	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	130	1.10e-07	0.3	17	7.78/66310	6.06/41492	
22	ACJ65514	CD910695	wh	Triticum aestivum	Globulin 3	97	1.30e-03	0.38	12	7.78/66350	6.1/20794	
233	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	78	0.017	0.22	14	7.78/66310	6.69/66285	
24	gi|215398472		viridiplantae	Triticum aestivum	Globulin 3B	77	0.025	0.15	8	7.36/56897	7.18/57462	
596	ACJ65514	CD916929	wh	Zea mays	Globulin	102	3.90e-04	0.36	9	7.78/66350	6.3/53094	
607	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	75	0.034	0.19	15	7.78/66310	6.27/67429	
68	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	289	1.40e-23	0.53	30	7.78/66310	7.24/67330	
746	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	94	5.00e-04	0.2	12	7.78/66310	5.6/44735	
79	gi|215398468		viridiplantae	Triticum aestivum	Globulin 3C	124	4.50e-07	0.41	15	9.15/38406	6.93/68535	
885	ACJ65514	CD909132	wh	Triticum aestivum	Globulin 3	128	9.90e-07	0.67	12	7.78/66350	5.59/37345	
908	ACJ65514	CJ651760	wh	Triticum aestivum	Globulin 3	92	3.70e-03	0.65	9	7.78/66350	5.75/38433	
979	ACJ65513	BQ579030	wh	Triticum aestivum	Globulin 3c	95	1.80e-03	0.45	11	9.15/38429	6.42/18274	
98	ACJ65514	CA711407	wh	Triticum aestivum	Globulin 3	81	0.049	0.4	10	7.78/66350	6.64/17638	
98b	ACJ65514	CA711407	wh	Triticum aestivum	Globulin 3	90	6.30e-03	0.42	10	7.78/66350	6.64/17638	
99	gi|215398472		viridiplantae	Triticum aestivum	Globulin 3B	77	0.022	0.16	9	7.36/56897	6.84/57019	
10	gi|215398470			Triticum aestivum	Globulin 3	88	1.90e-03	0.27	14	7.78/66310	7.76/57004	
782	ACJ65514	CA740959	wh	Triticum aestivum	Globulin 3	87	0.012	0.57	10	7.78/66350	7.08/41427	
787	ACJ65514	CD906551	wh	Triticum aestivum	Globulin 3	104	2.50e-04	0.38	8	7.78/66350	5.58/38089	
221	ABF95817	BJ289323	wh	Triticum aestivum	Cupin family protein, expressed	87	0.012	0.43	6	7.18/61439	5.48/33989	
771	gi|215398470		viridiplantae	Oryza sativa Japonica Group	Globulin 3	130	1.80e-03	0.06	2	7.78/66310	5.48/37502	GSSNLQVVCFEINAER +2 Carbamidomethyl (C), GSAFVVPPGHPVVEIASSR +2

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Coverage rate	Number of matched peptides	Experimental pI/MW (Da)	Theoratical pI/MW (Da)	MS/MS peptides sequence with valid score
932	gi|215398472		viridiplantae	Triticum aestivum	Globulin3B	174	5.80e-06	0.07	2	7.36/56897	6.03/79778	DQQDEGFVAGPEQQEQER +2, AKDQQDEGFVAGPEQQEQER +2
478	AAM33459	CD910595	wh	Triticum aestivum	Globulin-like protein	132	3.90e-07	0.54	13	6.79/52066	6.85/49377	
118	gi|215398470		viridiplantae	Triticum aestivum	Globulin 3	127	1.30e-03	0.05	2	7.78/66310	4.9/21318	GSSNLQVV**C**FEINAER +2 Carbamidomethyl (C), GSAFVVPPGHPVVEIASSR +2
46	AAM33459	BQ806323	wh	Triticum aestivum	Globulin-like protein	117	1.20e-05	0.64	12	6.79/52066	7.44/60265	
Lipid metabolism												
296	CAA57994	BQ805416	wh	Hordeum vulgare subsp. Vulgare	High molecular weight oleosin	85	0.019	0.29	5	9.69/18494	7.24/50560	
Metabolism of other amino acid												
1174*	gi|20067415		viridiplantae	Triticum aestivum	Glutathione transferase	299	9.80e-04	0.28	5	6.35/24984	5.84/27630	AE**M**LDILK +2 Oxidation (M), IPVLLHDGR +2, TLDGALGDKPFFGGDK +2, GlPYEYAEEDL**M**AGK +2 Oxidation (M), SLYSPDKVYDFIGILK +2
867*	AAD56395	BJ267219	wh	Triticum aestivum	Glutathione S-transferase	118	9.90e-06	0.46	9	5.79/23339	5.76/28254	
Biosynthesis of secondary metabolites												
205	CAA88322	BJ293013	wh	Hordeum vulgare	Aldose reductase	85	0.018	0.41	7	6.68/35906	6.54/30454	
595	P23901	BJ291837	wh	Hordeum vulgare	Aldose reductase	82	0.039	0.54	9	6.51/35807	6.4/38459	
993	gi|46946550		viridiplantae	Triticum aestivum	Polyphenol oxidase	83	0.006	0.3	11	5.23/46550	5.1/61240	
GENETIC INFORMATION AND PROCESSING												
Protein synthesis												
352	NP_001151139	CD917915	wh	Zea mays	Heat shock protein	104	2.50e-04	0.53	13	6/22886	5.31/28582	
1065	gi|213959111		viridiplantae	Oryza sativa, Japonica Group	HSP70	129	1.40e-07	0.37	21	5.30/71648	5.08/66802	
339	AAD46133	CD920118	wh	Triticum aestivum	Secretory protein	119	7.90e-06	0.52	12	9.32/24210	9.12/26528	
597	gi|157093720		viridiplantae	Triticum aestivum	Secretory protein cursor	74	0.041	0.34	6	9.32/24210	8.76/27157	
Protein Folding												
1482	gi|12056115		viridiplantae	Triticum turgidum subsp. durum	Protein disulfide isomerase	275	3.50e-22	0.55	26	4.99/56544	4.88/63895	
1403	gi|222446340		viridiplantae	Triticum aestivum	Protein disulfide isomerase	230	1.10e-17	0.45	23	4.96/56636	4.85/64504	
ENVIRONMENTAL INFORMATION AND PROCESSING												
Membrane transport												
1110	ACG48482	BJ291619	wh	Zea mays	Lipoprotein	80	0.063	0.42	7	5.87/26214	5.17/32645	
1334	ACG48482	CD910501	wh	Zea mays	Lipoprotein	94	2.80e-03	0.36	7	5.87/26214	5.17/33075	
951	ACG48482	CK206290	wh	Zea mays	Lipoprotein	99	7.70e-04	0.26	9	5.87/26214	5.28/32645	
STRESS DEFENSE												
Miscellanous stress proteins												
129	AAX99161	CD919878	wh	Hordeum vulgare subsp. Vulgare	2-alkenal reductase	124	2.50e-06	0.59	14	5.88/38257	5.26/43655	
1490	gi|119388733		viridiplantae	Triticum turgidum subsp. dicoccoides	Alcohol dehydrogenase	128	1.80e-07	0.51	18	6.28/41010	6.03/49737	
1538*	gi|1621627		viridiplantae	Triticum aestivum	Manganese superoxide dismutase	319	4.30e-04	0.36	5	7.9/25259	6/27062	GDASAVVHLQSAIK +2, NLKPISEGGGEAPHGK +3, FNGGGHVNHSIFWK +2, LSVETTPNQDPLVTK +2, ALEQLDAAVSKGDASAVVHLQSAIK +3

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Coverage rate	Number of matched peptides	Experimental pI/MW (Da)	Theoratical pI/MW (Da)	MS/MS peptides sequence with valid score
127	gi|20257409			Triticum aestivum	Thaumatin-like protein	227	2.60e-07	0.23	3	7.85/23582	7.29/29036	GQFEHN**C**PPTNYSK +2 Carbamidomethyl (C), LDPGQSWALN**M**PAGTAGAR +2 Oxidation (M), VSGQQPTTLAEYTLGQGGNK +2,
105	AAM15877	CN011817	wh	Triticum aestivum	Thaumatin-like protein	76	0.17	0.58	10	7.85/23598	7.58/28723	
169	gi|123975		viridiplantae	Triticum aestivum	Alpha-amylase/subtilisin inhibitor	180	1.10e-12	0.83	16	6.77/19621	6.67/21070	
983*	NP_001148437	BQ804151	wh	Zea mays	Peroxiredoxin-6	86	0.016	0.52	7	7.74/23819	4.84/22051	
1466*	gi|1654387		viridiplantae	Triticum aestivum	Manganese superoxide dismutase	108	1.80e-05	0.59	11	7.89/25283	5.89/27019	
236	gi|123975		viridiplantae	Triticum aestivum	Alpha-amylase/subtilisin inhibitor;	150	1.10e-09	0.8	13	6.77/19621	7.14/23092	
762	gi|123975		viridiplantae	Triticum aestivum	Endogenous alpha-amylase/subtilisin inhibitor	106	2.80e-05	0.73	11	6.77/19621	6.47/22465	
133*	ABF95269	BJ289536	wh	Oryza sativa, Japonica Group	Glyoxalase family protein, expressed	90	5.60e-03	0.46	10	9.44/23357	5.94/16231	
816*	gi|28192421		viridiplantae	Triticum aestivum	Dehydroascorbate reductase	110	1.10e-05	0.63	11	5.88/23343	5.55/28114	
201*	gi|34539782		viridiplantae	Triticum aestivum	1-Cys-peroxiredoxine	365	3.80e-05	0.42	5	6.08/23950	6.05/29440	AVDSLLTAAK +2, LLGIS**C**DDVQSHK +2 Carbamidomethyl (C), **M**FPQGFETADLPSK +2 Oxidation (M), PGITIGDTVPNLELDSTHGK +2, IHDYVGNGYVILFSHPGDFTPV**C**TTELAA**M**ANYAK +3 Carbamidomethyl (C), Oxidation (M)
Cell wall synthesis/related defence												
5	gi|62465514		viridiplantae	Triticum aestivum	Class II chitinase	84	4.30e-03	0.55	10	8.66/28200	8/30295	
8	gi|62465514		viridiplantae	Triticum aestivum	Class II chitinase	82	7.90e-03	0.37	8	8.66/28200	8.22/30460	
11	gi|62465514		viridiplantae	Triticum aestivum	Class II chitinase	132	7.10e-08	0.67	13	8.66/28200	7.71/29977	
25	gi|75262903		viridiplantae	Secale cereale	Basic endochitinase	77	0.024	0.31	7	8.82/28284	8.78/30797	
227	gi|51247633		viridiplantae	Triticum aestivum	Xylanase In Complex With Inhibitor (Xip-I)	157	2.20e-10	0.54	17	8.27/30266	7.55/33912	
14	gi|75262903		viridiplantae	Secale cereale	Endochitinase	78	0.017	0.35	10	8.82/28284	8.59/30891	
9	AAX83262	CD909727	wh	Triticum aestivum	Class II chitinase	80	0.069	0.49	6	8.66/28218	8.45/30938	
1348	gi|6175480		viridiplantae	Hordeum vulgare	Xylose isomerase	139	1.40e-08	0.43	16	5.31/53580	5.2/61938	
4	gi|51247633		viridiplantae	Triticum aestivum	Xylanase In Complex With Inhibitor (Xip-I)	172	7.10e-12	0.62	17	8.27/30266	7.99/32357	
27	gi|51247633		viridiplantae	Triticum aestivum	Xylanase In Complex With Inhibitor (Xip-I)	146	2.80e-09	0.59	17	8.27/30266	7.99/35038	
90	gi|51247633		viridiplantae	Triticum aestivum	Xylanase In Complex With Inhibitor (Xip-I)	159	1.40e-10	0.51	18	8.27/30266	7.98/36235	
1114	ACQ41883	CK206848	wh	Oryza sativa, Japonica Group	Germin-like protein 3	198	1.60e-03	0.13	3	6.5/23370	4.65/79388	ATEIGlVVK +2, VDFAPGGTNPPHIHPR +2, LDVTEFPGENTQGISMNR +2 Oxidation (M)
297	AJ422119		wh	Triticum aestivum	mRNA for xylanase inhibitor protein I	83	0.034	0.27	7	8.66/33192	7.05/25628	
67	gi|51247633		viridiplantae	Triticum aestivum	Xylanase In Complex With Inhibitor (Xip-I)	140	1.10e-08	0.47	16	8.27/30266	7.91/32605	
3	gi|31615809		viridiplantae	Triticum aestivum	Xylanase Inhibitor Protein (Xip-I) From Wheat	117	2.20e-06	0.5	15	8.27/30266	7.99/34007	
UNKNOWN / HYPOTHETICAL PROTEIN												
1478	EAZ13538	BQ842063	wh	Oryza sativa, Japonica Group	Hypothetical protein OsJ_03454	93	0.003	0.47	10	9.77/37776	6.65/37273	
253	EAZ13538	CD910590	wh	Oryza sativa Japonica Group	Hypothetical protein OsJ_03454	106	1.60e-04	0.52	11	9.77/37776	7.01/35798	
487	EEC76319	AL818892	wh	Oryza sativa Indica Group	Hypothetical protein	109	7.90e-05	0.7	7	6.99/52060	6.59/57040	
560	EEC76319	BQ246191	wh	Oryza sativa Indica Group	Hypothetical protein	108	9.90e-05	0.48	12	6.99/52060	6.65/56575	
481	EAY88907	CD925599	wh	Oryza sativa Indica Group	Hypothetical protein OsI_10386	193	3.30e-03	0.28	3	5.74/68245	4.91/50570	EGDVFVVPR +1, EIVALALGQK +1, VKEGDVFVVPR +2
620	XP_002465800	CD925401	wh	Sorghum bicolor	Hypothetical protein SORBIDRAFT_01g046000	139	1.80e-03	0.17	2	6.67/36503	8.68/55275	ETVQHTASETGR +2, TTPGGAEVAPGKEEESWTGWAK +2
369	XP_002468335	CD925597	wh	Sorghum bicolor	Hypothetical protein SORBIDRAFT_01g043980	197	0.002	0.28	3	5.74/68245	4.94/49809	EGDVFVVPR +2, EIVALALGQK +2, VKEGDVFVVPR +2

SuppInfo 5: Chloroplastic proteins identified by MALDI-TOF or MS/MS in PL's during grain development.

Spot ID	NCBI accession number	Wh-EST database accession number	Database	Taxonomy	Protein Name	Mascot Score	Expect Value	Couverage rate	Number of matched peptides	Experimental pI/MW (Da)	MS/MS peptides sequence with valid score
540	gi|4038721		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	99	1.20e-04	0.57	10	8.83/18533	
581	gi|544700		viridiplantae	Hordeum vulgare	Light-harvesting complex I	204	1.30e-04	0.1	3	8.13/24211	YPGGAFDPLGFSK +2, FKESEIYH**C**R +2 Carbamidomethyl (C), KYPGGAFDPLGFSK +2
612	ACG28457	BE591357	wh	zea mays	Chlorophyll a-b binding protein 8	80	0.06	0.46	6	8.94/28984	
448	Q7X9A6	CJ897844	wh	Triticum aestivum	Cytochrome b6-f complex iron-sulfur subunit.	91	0.0045	0.53	10	8.47/23726	
455	gi|61378618		viridiplantae	Triticum monococcum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	152	6.10e-10	0.36	20	6.04/52716	
23	gi|134290407		viridiplantae	Triticum aestivum	Putative oxygen-evolving complex precursor	78	0.014	0.41	9	9.72/21126	
26	gi|14017580		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	204	3.80e-15	0.44	25	6.22/52817	
35	gi|11990897		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	145	3.10e-09	0.66	15	8.8/19448	
41	gi|18958		viridiplantae	Hordeum vulgare subsp. vulgare	Chlorophyll a/b binding (CAB) protein of photosystem II (PSII)	82	0.007	0.3	7	5.33/30725	
50	gi|169666638		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	176	2.40e-12	0.39	22	6.22/52817	
1529	AAB82711	CJ906759	wh	x Tritordeum sp.	Glycine decarboxylase P subunit	88	0.0093	0.33	7	6.32/111020	
16	gi|13310805		viridiplantae	Eriachne triodioides	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	81	0.0075	0.28	14	6.22/46259	
18	gi|14017580		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	255	3.10e-20	0.47	27	6.22/52817	
19	gi|147945622		viridiplantae	Leymus chinensis	Chloroplast oxygen-evolving enhancer protein 1	112	6.10e-06	0.44	13	6.08/34490	
21	gi|119710748		viridiplantae	Thysananthus anguiformis	Ribulose 1,5-bisphosphate carboxylase/oxygenase	75	0.031	0.3	7	8.86/34295	
103	NP_00114843 9	CJ930673	wh	Zea mays	Chlorophyll a-b binding protein 2	102	3.90e-04	0.33	7	5.46/27945	
141	gi|14017580		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	106	2.40e-05	0.31	17	6.22/52817	
153	gi|11990897		viridiplantae	Triticum aestivum	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	113	4.80e-06	0.66	14	8.8/19448	
796	gi|57283816		viridiplantae	Dactylis Glomerata	Ribulose bisphosphate carboxylase large chain	121	8.90e-07	0.28	17	6.04/51203	
733	gi|54303886		viridiplantae	Molinia caerulea	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	77	0.018	0.24	12	6.13/46678	
120	gi|31087891		viridiplantae	Hordeum patagonicum subsp. santacrucense	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	91	8.20e-04	0.24	15	5.95/52739	
12	gi|134102		viridiplantae	Triticum aestivum	RuBisCO	195	3.10e-14	0.55	21	4.83/57485	
277	gi|861010		viridiplantae	Hordeum vulgare subsp. vulgare	2-Cys peroxiredoxin	106	2.40e-05	0.43	6	5.48/23284	
1193	gi|1174745		viridiplantae	Secale cereale	Triosephosphate isomerase	74	0.037	0.42	9	6/31613	
667	gi|1174745		viridiplantae	Secale cereale	Triosephosphate isomerase	99	1.20e-04	0.48	11	6/31613	
679	AAF15312	CJ721551	wh	Oryza sativa	Chloroplast translational elongation factor Tu	82	0.037	0.43	7	6.05/50355	
328	ACM78035	DV478664	poa	Triticum aestivum	Chloroplast fructose-bisphosphate aldolase	88	0.035	0.37	8	5.94/42015	

ANNEXE 2: Supplemental Information data of Article 3 (Chapter 4); An attempt to integrate proteome and transcriptome data: application to carbohydrate metabolism of wheat grain development.

SuppInfo Table 1: Endosperm proteins and their corresponding Transcripts obtained after tblastn for CM proteins. Spot ID, Accession, protein identifier in NCBI; Ta#S id, transcript identifier in NCBI unigene; Ta. Id, transcript id in NCBI unigen; Accession after BlastP, protein id after BlastP; EC number, enzyme number; BINCODE, Mapman id; protein name and predicted sub-cellular localization of protein

Spot ID	Accession	Ta#S id	Accession after BlastP	EC number	BINCODE	Protein name	Predicted sub-cellular localization
344	gi183211902	Ta#S43837843		2.4.1.1	2.2.2.2	Plastid alpha-1,4-glucan phosphorylase	P
348	gi183211902	Ta#S43837843		2.4.1.1	2.2.2.2	Plastid alpha-1,4-glucan phosphorylase	P
400	gi75225211	Ta#S13111784		4.2.1.3	8.1.3	Putative aconitate hydratase	cyto
401	gi75225211	Ta#S13111784		4.2.1.3	8.1.3	Putative aconitate hydratase	cyto
408	gi75225211	Ta#S13111784		4.2.1.3	8.1.3	Putative aconitate hydratase	cyto
547	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
554	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
555	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
561	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
570	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
571	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
573	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
576	gi3393044	Ta#S12922949		2.4.1.13	2.2.1.5	Sucrose synthase type 2	cyto
604	gi227483057	Ta#S17888448	gi28190676	2.2.1.1	7.2.1	Putative transketolase	P
615	gi227483057	Ta#S17888448	gi28190676	2.2.1.1	7.2.1	Putative transketolase	P
774	gi115476012	Ta#S37770968	Q84QT9	2.7.1.90	4.1.6	Putative pyrophosphate-dependent phosphofructokinase alpha subunit	cyto
779	gi162957175	Ta#S50084340		1.1.1.40	8.2.10	NADP-dependent malic enzyme 1	P
803	gi259708196	Ta#S24623093	F2CX32	2.7.1.40	4.1.15	Pyruvate kinase	cyto
810	gi115448277	Ta#S26026583	Q6ZFT9	2.7.1.90	4.1.6	Putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain	cyto
876	gi298545815	Ta#S15116139	gi4588609	2.4.1.242	2.1.2.2	Granule bound starch synthase precursor	P
877	gi18449343	Ta#S12923502		1.5.1.12	13.2.2.2	Putative aldehyde dehydrogenase WIS1	M
882	gi38349539	Ta#S17989497		3.2.1.2	2.2.2.1	Beta-amylase 1	Extr
883	gi38349539	Ta#S17989497		3.2.1.2	2.2.2.1	Beta-amylase 1	Extr
889	gi38349539	Ta#S17989497		3.2.1.2	2.2.2.1	Beta-amylase 1	Extr
890	gi91694277	Ta#S32418945		5.3.1.9	4.1.4	Glucose-6-phosphate isomerase	cyto
918	gi91694277	Ta#S32418945		5.3.1.9	4.1.4	Glucose-6-phosphate isomerase	cyto
920	gi91694277	Ta#S32418945		5.3.1.9	4.1.4	Glucose-6-phosphate isomerase	cyto
1020*	gi242089615	Ta#S16058021	C5Z0L0	1.8.1.4	8.1.1.3	Dihydrolipoyl dehydrogenase	P
1113*	gi218401053	Ta#S17890115	gi108708038	4.2.1.2	8.1.8	Fumarate hydratase 1	M
1116	gi218401053	Ta#S17890115	gi108708038	4.2.1.2	8.1.8	Fumarate hydratase 1	M
1124	gi226506764	Ta#S13112438	B4FRC9	2.2.1.2	7.2.2	Transaldolase 2	P
1183	gi242064362	Ta#S17986739		2.3.3.1	8.1.2	Hypothetical protein SORBIDRAFT_04g006440	M

1200	gi288814543	Ta#S17988960		3.2.1.1	2.1.2.4		Alpha amylase	Extr
1235	gi298549023	Ta#S13098196	gi226499486	1.1.1.42	8.2.4		Isocitrate dehydrogenase2	cyto
1259*	gi242095836	Ta#S12922033	gi195640660	1.2.1.2		25	Formate dehydrogenase 1	M
1294	gi115447367	Ta#S17988642	Q6K9N6				Succinyl-CoA synthetase beta chain	M
1356	gi226316439	Ta#S24623187		4.1.2.13	4.1.11		Fructose-bisphosphate aldolase	cyto
1370	gi125560179	Ta#S32570711	A2YRB6				Transaldolase family protein	P
1404*	gi125561648	Ta#S37941675	A2YVI5	1.1.1.37	8.1.9		Malate dehydrogenase	P
1413	gi32400859	Ta#S13208529		4.1.2.13	4.1.11		Aldolase	P
1415*	gi125561648	Ta#S37941675	A2YVI5	1.1.1.37	8.1.9		Malate dehydrogenase	P
1418	gi223018643	Ta#S17888674		4.1.2.13	4.1.11		Fructose-bisphosphate aldolase	P
1419	gi223018643	Ta#S17888674		4.1.2.13	4.1.11		Fructose-bisphosphate aldolase	P
1427	gi32400859	Ta#S13208529		4.1.2.13	4.1.11		Aldolase	P
1458	gi34485587	Ta#S43837843		2.4.1.1	2.2.2.2		Plastidic alpha 1,4-glucan phosphorylase 3	cyto
2126	gi259703023	Ta#S17879701	gi11124572	5.3.1.1	4.1.8		Triosephosphat-isomerase	cyto
2549	gi210063883	Ta#S12922947		2.7.7.27	2.1.2.1		Putative glucose-1-phosphate adenylyltransferase large subunit 1	P
1255b	gi298549023	Ta#S13098196	gi226499486	1.1.1.42	8.2.4		Isocitrate dehydrogenase2	cyto
883a	gi38349539	Ta#S17989497		3.2.1.2	2.2.2.1		Beta-amylase 1	Extr
384	gi257665965	Ta#S17897781	gi3341490	4.1.1.31	4.1.16		Phospoenolpyruvate carboxylase	cyto
539	gi11037530	Ta#S12923266		2.4.1.18	2.1.2.3		Starch branching enyzyme 1	P
540	gi11037530	Ta#S12923266		2.4.1.18	2.1.2.3		Starch branching enyzyme 1	P
546	gi11037530	Ta#S12923266		2.4.1.18	2.1.2.3		Starch branching enzyme 1	P
552	gi401138	Ta#S12923095		2.4.1.13	2.2.1.5		Sucrose synthase 1	M
560	gi401138	Ta#S12923095		2.4.1.13	2.2.1.5		Sucrose synthase 1	M
568	gi401138	Ta#S12923095		2.4.1.13	2.2.1.5		Sucrose synthase 1	M
569	gi401138	Ta#S12923095		2.4.1.13	2.2.1.5		Sucrose synthase 1	M
735	gi18076790	Ta#S16058207		5.4.2.2	4.1.3		Phosphoglucomutase	cyto
759	gi18076790	Ta#S16058207		5.4.2.2	4.1.3		Phosphoglucomutase	cyto
769	gi32400802	Ta#S13132127		5.4.2.1	4.1.13		Phosphoglycerate mutase	cyto
775	gi32400802	Ta#S13132127		5.4.2.1	4.1.13		Phosphoglycerate mutase	cyto
775	gi32400802	Ta#S32623200		5.4.2.1	4.1.13		Phosphoglycerate mutase	cyto
778	gi32400802	Ta#S13132127		5.4.2.1	4.1.13		Phosphoglycerate mutase	cyto
798*	gi158701881	Ta#S41658089		1.1.1.40	8.2.10		NADP-dependent malic enzyme	cyto
892	gi296523708	Ta#S13129182	gi162457852	4.1.1.1		5,2	Pyruvate decarboxylase isozyme 3	cyto
892	gi296523708	Ta#S17985756	gi162457852	4.1.1.1		5,2	Pyruvate decarboxylase isozyme 3	cyto
910	gi195619804	Ta#S17897418		4.2.1.11	4.1.14		Enolase	cyto
910	gi195619804	Ta#S32671937		4.2.1.11	4.1.14		Enolase	cyto
936	gi195619804	Ta#S17897418		4.2.1.11	4.1.14		Enolase	cyto
936	gi195619804	Ta#S32671937		4.2.1.11	4.1.14		Enolase	cyto
943	gi115455455	Ta#S16057998	gi195623986	1.1.1.22	10.1.4		UDP-glucose 6-dehydrogenase	ER
956	gi1707923	Ta#S12922947		2.7.7.27	2.1.2.1		Glucose-1-phosphate adenylyltransferase large subunit 1	P

959	gi195619804	Ta#S32671937		4.2.1.11	4.1.14		Enolase	cyto
965	gi193073259	Ta#S32529534		3.2.1.21	10.6.1		Beta-glucosidase	ER
966	gi1707923	Ta#S12922947		2.7.7.27	2.1.2.1		Glucose-1-phosphate adenylyltransferase large subunit 1	P
967	gi1707930	Ta#S12922947		2.7.7.27	2.1.2.1		Glucose-1-phosphate adenylyltransferase large subunit	P
968	gi193073259	Ta#S32529534		3.2.1.21	10.6.1		Beta-glucosidase	ER
974	gi193073259	Ta#S32529534		3.2.1.21	10.6.1		Beta-glucosidase	ER
974	gi193073259	Ta#S32617924		3.2.1.21	10.6.1		Beta-glucosidase	ER
975	gi193073259	Ta#S32529534		3.2.1.21	10.6.1		Beta-glucosidase	ER
994*	gi291047856	Ta#S17985962	gi20530127	1.2.1.3	5.10.		Mitochondrial aldehyde dehydrogenase	M
998	gi20127139	Ta#S12922948		2.7.7.27	2.1.2.1		Small subunit ADP glucose pyrophosphorylase	P
1004	gi20127139	Ta#S12922948		2.7.7.27	2.1.2.1		Small subunit ADP glucose pyrophosphorylase	P
1012	gi6136111	Ta#S13044452		2.7.7.9	4.1.2		UTP--glucose-1-phosphate uridylyltransferase	cyto
1012	gi6136111	Ta#S16058223		2.7.7.9	4.1.2		UTP--glucose-1-phosphate uridylyltransferase	cyto
1033	gi6136111	Ta#S13044452		2.7.7.9	4.1.2		UTP--glucose-1-phosphate uridylyltransferase	cyto
1044	gi1707930	Ta#S12922947		2.7.7.27	2.1.2.1		Glucose-1-phosphate adenylyltransferase large subunit	P
1067	gi1707930	Ta#S12922947		2.7.7.27	2.1.2.1		Glucose-1-phosphate adenylyltransferase large subunit	P
1077	gi148508784	Ta#S12942128		1.2.1.12	4.1.9		Glyceraldehyde-3-phosphate dehydrogenase	M
1077	gi148508784	Ta#S17887997		1.2.1.12	4.1.9		Glyceraldehyde-3-phosphate dehydrogenase	M
1077	gi148508784	Ta#S18009604		1.2.1.12	4.1.9		Glyceraldehyde-3-phosphate dehydrogenase	M
1106	gi224021585	Ta#S12922948		2.7.7.27	2.1.2.1		Plastid ADP-glucose pyrophosphorylase small subunit	P
1154	gi298541521	Ta#S12923000	gi129916	2.7.2.3	4.1.12		Phosphoglycerate kinase, cytosolic	cyto
1191*	gi242095836	Ta#S12922033	gi195640660	1.2.1.2		25	Formate dehydrogenase 1	M
1191*	gi242095836	Ta#S17888699	gi195640660	1.2.1.2		25	Formate dehydrogenase 1	M
1202*	gi242095836	Ta#S12922033	gi195640660	1.2.1.2		25	Formate dehydrogenase 1	M
1202*	gi242095836	Ta#S17888699	gi195640660	1.2.1.2		25	Formate dehydrogenase 1	M
1256*	gi242095836	Ta#S12922033	gi195640660	1.2.1.2		25	Formate dehydrogenase 1	M
1256*	gi242095836	Ta#S17888699	gi195640660	1.2.1.2		25	Formate dehydrogenase 1	M
1286	gi115447367	Ta#S17984871	Q6K9N6	6.2.1.5	8.1.6		Succinyl-CoA synthetase beta chain	M
1295	gi253783729	Ta#S13110859		1.2.1.12	4.1.9		Glyceraldehyde-3-phosphate dehydrogenase	cyto
1305	gi28172909	Ta#S12923000		2.7.2.3	4.1.12		Cytosolic 3-phosphoglycerate kinase	cyto
1318	gi253783729	Ta#S12942128		1.2.1.12	4.1.9		Glyceraldehyde-3-phosphate dehydrogenase	cyto
1318	gi253783729	Ta#S16245770		1.2.1.12	4.1.9		Glyceraldehyde-3-phosphate dehydrogenase	cyto
1324	gi298541521	Ta#S12923000	gi129916	2.7.2.3	4.1.12		Phosphoglycerate kinase, cytosolic	cyto
1332	gi226316439	Ta#S24623187		4.1.2.13	4.1.11		Fructose-bisphosphate aldolase	cyto
1333	gi3646373	Ta#S12922855		2.4.1.-	10.5.5		RGP1 protein	cyto
1347	gi4158232	Ta#S12922855		2.4.1.-	10.5.5		Reversibly glycosylated polypeptide	cyto
1352*	gi229358240	Ta#S37366469		1.1.1.37	8.1.9		Cytosolic malate dehydrogenase	cyto
1355	gi226316439	Ta#S24623187		4.1.2.13	4.1.11		Fructose-bisphosphate aldolase	cyto
1366	gi226316439	Ta#S24623187		4.1.2.13	4.1.11		Fructose-bisphosphate aldolase	cyto
1372	gi148508784	Ta#S37810935		1.2.1.12	4.1.9		Glyceraldehyde-3-phosphate dehydrogenase	M

1374*	gi229358240	Ta#S37366469		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
1376*	gi229358240	Ta#S37366469		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
1380	gi148508784	Ta#S37810935		1.2.1.12	4.1.9	Glyceraldehyde-3-phosphate dehydrogenase	M
1386	gi148508784	Ta#S37810935		1.2.1.12	4.1.9	Glyceraldehyde-3-phosphate dehydrogenase	M
1393*	gi229358240	Ta#S37366469		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
1399*	gi229358240	Ta#S37366469		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
1436*	gi115465579	Ta#S13111477	Q6F361	1.1.1.37	8.1.9	Malate dehydrogenase	M
1436*	gi115465579	Ta#S16058229	Q6F361	1.1.1.37	8.1.9	Malate dehydrogenase	M
1444*	gi115465579	Ta#S13111477	Q6F361	1.1.1.37	8.1.9	Malate dehydrogenase	M
1444*	gi115465579	Ta#S16058229	Q6F361	1.1.1.37	8.1.9	Malate dehydrogenase	M
1455*	gi195628708	Ta#S13111477		1.1.1.37	8.1.9	Malate dehydrogenase	M
1483	gi226316439	Ta#S24623187		4.1.2.13	4.1.11	Fructose-bisphosphate aldolase	cyto
1489*	gi32400786	Ta#S17988504		1.1.1.224	3.5	NADPH-dependent mannose 6-phosphate reductase	perox
1543	gi7579064	Ta#S12942128		1.2.1.12	4.1.9	Cytosolic glyceraldehyde-3-phosphate dehydrogenase	cyto
1543	gi7579064	Ta#S18009604		1.2.1.12	4.1.9	Cytosolic glyceraldehyde-3-phosphate dehydrogenase	cyto
1543	gi7579064	Ta#S37810935		1.2.1.12	4.1.9	Cytosolic glyceraldehyde-3-phosphate dehydrogenase	cyto
1784	gi257355184	Ta#S18005976	O48556	3.6.1.1	23.4.99	Soluble inorganic pyrophosphatase	cyto
1798	gi259662377	Ta#S37826290	gi226529672	5.3.1.1	4.1.8	Triosephosphate isomerase, cytosolic	cyto
1855	gi259703023	Ta#S18010579	gi11124572	5.3.1.1	4.1.8	Triosephosphat-isomerase	cyto
1863	gi290875537	Ta#S12906559		4.2.1.1	8.3	Putative carbonic anhydrase	P
1908	gi122171022	Ta#S18005976		3.6.1.1	23.4.99	Soluble inorganic pyrophosphatase	cysk
2635	gi18181983	Ta#S17989286		5.5.1.4	3.4.3	Myo-inositol-1-phosphate synthase	cyto
2650	gi148508784	Ta#S37810935		1.2.1.12	4.1.9	Glyceraldehyde-3-phosphate dehydrogenase	M
2651	gi148508784	Ta#S37810935		1.2.1.12	4.1.9	Glyceraldehyde-3-phosphate dehydrogenase	M
2654*	gi229358240	Ta#S13046790		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
2657*	gi229358240	Ta#S13046790		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
2660*	gi229358240	Ta#S13046790		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
2660*	gi229358240	Ta#S37366469		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
2664*	gi229358240	Ta#S37366469		1.1.1.37	8.1.9	Cytosolic malate dehydrogenase	cyto
1330b	gi298541521	Ta#S12923000	gi129916	2.7.2.3	4.1.12	Phosphoglycerate kinase, cytosolic	cyto
1365a	gi226316439	Ta#S24623187		4.1.2.13	4.1.11	Fructose-bisphosphate aldolase	cyto
1365b	gi226316439	Ta#S24623187		4.1.2.13	4.1.11	Fructose-bisphosphate aldolase	cyto
1366a	gi226316439	Ta#S24623187		4.1.2.13	4.1.11	Fructose-bisphosphate aldolase	cyto
718a	gi20259685	Ta#S12989027		3.2.1.21	10.6.1	Beta-D-glucan exohydrolase	ER
718a	gi20259685	Ta#S22389515		3.2.1.21	10.6.1	Beta-D-glucan exohydrolase	ER
885b*	gi218196143	Ta#S15902783	B8AYE1	1.8.1.4	8.1.1.3	Dihydrolipoamide dehydrogenase	M
885b*	gi218196143	Ta#S17984338	B8AYE1	1.8.1.4	8.1.1.3	Dihydrolipoamide dehydrogenase	M
326	gi222628767	no hit		1.2.4.2	8.1.5	Hypothetical protein OsJ_14584	M
370	gi222628767	no hit		1.2.4.2	8.1.5	Hypothetical protein OsJ_14584	M
411	gi92429669	no hit		4.2.1.3	8.1.3	Putative aconitate hydratase 1	P
823	gi115467370	no hit	gi15221156	2.7.1.90	4.1.6	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
824	gi115467370	no hit	gi15221156	2.7.1.90	4.1.6	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
828	gi115467370	no hit	gi15221156	2.7.1.90	4.1.6	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
836	gi32400764	no hit		3.2.1.2	2.2.2.1	Beta-amylase	Extr
837	gi32400764	no hit		3.2.1.2	2.2.2.1	Beta-amylase	Extr
848	gi115467370	no hit	gi15221156	2.7.1.90	4.1.6	Pyrophosphate--fructose-6-phosphate 1-phosphotransferase	P
898	gi32400764	no hit		3.2.1.2	2.2.2.1	Beta amylase	Extr
909	gi38349539	no hit		3.2.1.2	2.2.2.1	Beta-amylase 1	Extr
1072	gi226532024	no hit		2.3.1.61	35.1*	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	M
1310	gi2204226	no hit		3.2.1.22	3.8.2	Alpha-galactosidase	nucl
1472	gi32400764	no hit		3.2.1.2	2.2.2.1	Beta amylase	Extr
1578	gi46805452	no hit		3.6.1.1	23.4.99	Putative inorganic pyrophosphatase	P

SuppInfo Table 2: Sequence alignment results when 1 protein matched to more than 1 transcript, using ClustalW2. Transcripts are first translated in their corresponding proteins and then the obtained translated sequence was used to align against corresponding protein sequence.

Match	Accession	Corresponding Translated TR	Score
1	gi|115465579|ref|NP_001056389.	gnl|UG|Ta#S13111477_+2_pep1	94.937
1	gi|115465579|ref|NP_001056389.	gnl|UG|Ta#S16058229_+3_pep1	91.636
2	gi|193073259|gb|ACF07998.1|	gnl|UG|Ta#S32617924_-1_pep1	93.694
2	gi|193073259|gb|ACF07998.1|	gnl|UG|Ta#S32529534_+1_pep1	40.483
3	gi|195619804|gb|ACG31732.1|	gnl|UG|Ta#S17897418_-3_pep1	93.077
3	gi|195619804|gb|ACG31732.1|	gnl|UG|Ta#S32671937_+1_pep1	47.558
4	gi|20259685|gb|AAM13694.1|	gnl|UG|Ta#S12989027_+1_pep1	100.000
4	gi|20259685|gb|AAM13694.1|	gnl|UG|Ta#S22389515_+1_pep1	94.387
5	gi|218196143|gb|EEC78570.1|	gnl|UG|Ta#S15902783_-2_pep1	100.000
5	gi|218196143|gb|EEC78570.1|	gnl|UG|Ta#S17984338_-3_pep1	64.458
5	gi|218196143|gb|EEC78570.1|	gnl|UG|Ta#S17984338_-3_pep2	50.820
6	gi|229358240|gb|ACQ57333.1|	gnl|UG|Ta#S37366469_+3_pep1	99.700
6	gi|229358240|gb|ACQ57333.1|	gnl|UG|Ta#S13046790_+1_pep1	90.598
7	gnl|UG|Ta#S12922033_+1_pep1	gi|242095836|ref|XP_002438408.	86.469
7	gnl|UG|Ta#S17888699_+1_pep1	gi|242095836|ref|XP_002438408.	83.636
8	gi|259703023|emb|CBG16603.1|	gnl|UG|Ta#S18010579_-1_pep2	79.470
8	gi|259703023|emb|CBG16603.1|	gnl|UG|Ta#S17879701_-1_pep1	64.706
8	gi|259703023|emb|CBG16603.1|	gnl|UG|Ta#S17879701_-1_pep2	47.723
8	gi|259703023|emb|CBG16603.1|	gnl|UG|Ta#S18010579_-1_pep1	36.055
9	gi|296523708|emb|CBM36708.1|	gnl|UG|Ta#S13129182_+1_pep1	100.000
9	gi|296523708|emb|CBM36708.1|	gnl|UG|Ta#S17985756_-3_pep1	93.426
10	gi|32400802|gb|AAP80633.1|AF47	gnl|UG|Ta#S13132127_+1_pep1	96.970
10	gi|32400802|gb|AAP80633.1|AF47	gnl|UG|Ta#S32623200_+1_pep1	89.578
11	gi|6136111|sp|Q43772.1|UGPA_HO	gnl|UG|Ta#S16058223_+3_pep1	98.943
11	gi|6136111|sp|Q43772.1|UGPA_HO	gnl|UG|Ta#S13044452_+1_pep1	80.271
12	gi|253783729|emb|CAZ76054.1|	gnl|UG|Ta#S12942128_-2_pep1	76.875
12	gi|253783729|emb|CAZ76054.1|	gnl|UG|Ta#S13110859_+1_pep1	30.524
12	gi|253783729|emb|CAZ76054.1|	gnl|UG|Ta#S16245770_+3_pep1	32.319
13	gi|7579064|gb|AAF64241.1|	gnl|UG|Ta#S12942128_-2_pep1	61.466
13	gi|7579064|gb|AAF64241.1|	gnl|UG|Ta#S18009604_-3_pep2	54.717
13	gi|7579064|gb|AAF64241.1|	gnl|UG|Ta#S37810935_+3_pep1	8.125
13	gi|7579064|gb|AAF64241.1|	gnl|UG|Ta#S18009604_-3_pep1	8.462
14	gi|148508784|gb|ABQ81648.1|	gnl|UG|Ta#S37810935_+3_pep1	90.972
14	gi|148508784|gb|ABQ81648.1|	gnl|UG|Ta#S17887997_-3_pep1	37.204
14	gi|148508784|gb|ABQ81648.1|	gnl|UG|Ta#S18009604_-3_pep2	34.713
14	gi|148508784|gb|ABQ81648.1|	gnl|UG|Ta#S12942128_-2_pep1	36.635
14	gi|148508784|gb|ABQ81648.1|	gnl|UG|Ta#S18009604_-3_pep1	32.985
14	gi|148508784|gb|ABQ81648.1|	gnl|UG|Ta#S17887997_-3_pep2	28.687

SuppInfo Table 3: Peripheral layers, carbohydrate metabolism proteins obtained after 2DE-MALDI-TOF. Spot ID, Accession, protein identifier in NCBI; EC number, emzyme number; Accession before BlastP, protein id obtained after Mascot search before BlastP.

Spot ID	Accession	EC number	Accession before BlastP	Protein name
1193	gi|1174745	5.3.1.1		Triosephosphate isomerase
456	NP_001105211	2.7.1.4	CK207916	Fructokinase 2
140	gi|19106	2.4.1.13		Sucrose synthase
667	gi|1174745	5.3.1.1		Triosephosphate isomerase
1198	gi|148508784	1.2.1.12		Glyceraldehyde-3-phosphate dehydrogenase
910	gi|219363507	1.1.1.42		Hypothetical protein
725	BAA22419	2.7.9.1	gi|218196777	Orthophosphate dikinase
1315	CAZ76054	1.2.1.12	gi|120680	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic
1504	CAC85913	5.4.2.2	gi|18076790	PhosphoGlucomutase
1401	ABB46862	4.2.1.11	gi|222624569	Enolase, putative, expressed
1464	CAA74160	3.2.1.22	CJ787667	Alpha-galactosidase
689	CAZ76054	1.2.1.12		Glyceraldehyde- 3-Phosphate Dehydrogenase
460	gi|8918504	1.1.1.49		Glucose-6-phosphate dehydrogenase
513	gi|8272480	4.1.2.13		Fructose 1,6-bisphosphate aldolase precursor
992	gi|37928995	1.1.1.37		Cytosolic malate dehydrogenase
147	gi|2507469	5.3.1.1		Triosephosphate isomerase
186	gi|8272480	4.1.2.13		Fructose 1,6-bisphosphate aldolase precursor
328	ACM78035	4.1.2.13	DV478664	Chloroplast fructose-bisphosphate aldolase
463	gi|7431022	1.1.1.-		Glucose and ribitol dehydrogenase
466	gi|7431022	1.1.1.-		Glucose and ribitol dehydrogenase
516	gi|113595	1.1.1.21		Aldose reductase
585	T06212	1.1.1.-	BQ805036	Glucose and ribitol dehydrogenase homolog
1195	CAA74160	3.2.1.22	CJ787667	Alpha-galactosidase.
1252	NP_001105292	1.1.1.40	BJ294714	NADP-dependent malic enzyme
459	gi|7431022	1.1.1.-		Glucose and ribitol dehydrogenase homolog
1291	gi|19851522	4.1.1.1		Pyruvate decarboxylase
626	gi|108708038	4.2.1.2		Fumarate hydratase 1, mitochondrial precursor, putative, expressed