Réponse du grain de blé à la nutrition azotée et soufrée : Étude intégrative des mécanismes moléculaires mis en jeu au cours du développement du grain par des analyses -omiques

Titouan Bonnot

Réponse du grain de blé à la nutrition azotée et soufrée : étude intégrative des mécanismes moléculaires mis en jeu au cours du développement du grain par des analyses -omiques

Titouan Bonnot

UNIVERSITE BLAISE PASCAL **UNIVERSITE D'AUVERGNE**
N°DU : 2767 ANNEE : 2016

ECOLE DOCTORALE SCIENCES DE LA VIE, SANTE, AGRONOMIE, ENVIRONNEMENT
N° d'ordre : 706

These :

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

DOCTEUR D'UNIVERSITE

(SPECIALITE : PHYSIOLOGIE ET GENETIQUE MOLECULAIRES)

soutenue le : 9 décembre 2016

TITOUAN BONNOT

Réponse du grain de blé à la nutrition azotée et soufrée : Etude intégrative des mécanismes moléculaires mis en jeu au cours du développement du grain par des analyses -omiques

TATOUT Christophe	Professeur Université Blaise Pascal, Clermont-Ferrand	**Président**
BUITINK Julia	Directeur de Recherches INRA, Angers	**Rapporteur**
RAJJOU Loïc	Maître de Conférences AgroParisTech, Paris	**Rapporteur**
GALLARDO Karine	Directeur de Recherches INRA, Dijon	**Examinateur**
ZIVY Michel	Directeur de Recherches CNRS, Gif-sur-Yvette	**Examinateur**
BANCEL Emmanuelle	Ingénieur d'Etudes INRA, Clermont-Ferrand	**Co-encadrante**
MARTRE Pierre	Directeur de Recherches INRA, Montpellier	**Co-encadrant**
RAVEL Catherine	Ingénieur de Recherches INRA, Clermont-Ferrand	**Directrice de thèse**

UMR 1095 INRA / UBP « Génétique, Diversité et Ecophysiologie des Céréales »
5, chemin de Beaulieu - 63039 Clermont-Ferrand Cedex 02 – France

Réponse du grain de blé à la nutrition azotée et soufrée : Etude intégrative des mécanismes moléculaires mis en jeu au cours du développement du grain par des analyses -omiques

RÉSUMÉ

L'augmentation des rendements est un enjeu majeur chez les céréales. Dans cet objectif, il est nécessaire de maintenir la qualité du grain de blé, qui est principalement déterminée par sa teneur et sa composition en protéines de réserve. En effet, une forte relation négative existe entre le rendement et la teneur en protéines. Par ailleurs, la qualité du grain est fortement influencée par la disponibilité en azote et en soufre dans le sol. La limitation des apports d'intrants azotés à la culture et la carence en soufre récemment observée dans les sols représentent ainsi des difficultés supplémentaires pour maitriser cette qualité. Une meilleure connaissance des mécanismes moléculaires impliqués dans le contrôle du développement du grain et la mise en place de ses réserves protéiques en réponse à la nutrition azotée et soufrée est donc primordiale. L'objectif de cette thèse a ainsi été d'apporter de nouveaux éléments à la compréhension de ces processus de régulation, aujourd'hui peu connus. Pour cela, les approches -omiques sont apparues comme une stratégie de choix pour identifier les acteurs moléculaires mis en jeu. Le protéome nucléaire a été une cible importante dans les travaux menés. L'étude de ces protéines nucléaires a révélé certains régulateurs transcriptionnels qui pourraient être impliqués dans le contrôle de la mise en place des réserves du grain. Dans une approche combinant des données de protéomique, transcriptomique et métabolomique, une vision intégrative de la réponse du grain à la nutrition azotée et soufrée a été obtenue. L'importance d'un apport de soufre dans le contrôle de la balance azote/soufre du grain, déterminante pour la composition du grain en protéines de réserve, a été clairement vérifiée. Parmi les changements observés au niveau du métabolisme cellulaire, certains des gènes affectés par la modification de cette balance pourraient orchestrer l'ajustement de la composition du grain face à des situations de carences nutritionnelles. Ces nouvelles connaissances devraient permettre de mieux maitriser la qualité du grain de blé dans un contexte d'agriculture durable.

Mots clés : Blé, grain, protéines de réserve, azote, soufre, protéines nucléaires, données omiques, réseaux biologiques

Wheat grain response to nitrogen and sulfur supply : Integrative study of molecular mechanisms involved during the grain development using -omics analyses

ABSTRACT

Improving the yield potential of cereals represents a major challenge. In this context, wheat grain quality has to be maintained. Indeed, grain quality is mainly determined by the content and the composition of storage proteins, but there is a strongly negative correlation between yield and grain protein concentration. In addition, grain quality is strongly influenced by the availability of nitrogen and sulfur in soils. Nowadays, the limitation of nitrogen inputs, and also the sulfur deficiency recently observed in soils represent major difficulties to control the quality. Therefore, understanding of molecular mechanisms controlling grain development and accumulation of storage proteins in response to nitrogen and sulfur supply is a major issue. The objective of this thesis was to create knowledge on the comprehension of these regulatory mechanisms. For this purpose, the best strategy to identify molecular actors involved in these processes consisted of -omics approaches. In our studies, the nuclear proteome was an important target. Among these proteins, we revealed some transcriptional regulators likely to be involved in the control of the accumulation of grain storage compounds. Using an approach combining proteomic, transcriptomic and metabolomic data, the characterization of the integrative grain response to the nitrogen and sulfur supply was obtained. Besides, our studies clearly confirmed the major influence of sulfur in the control of the nitrogen/sulfur balance that determines the grain storage protein composition. Among the changes observed in the cell metabolism, some genes were disturbed by the modification of this balance. Thus these genes could coordinate the adjustment of grain composition in response to nutritional deficiencies. These new results contribute in facing the challenge of maintaining wheat grain quality with sustainable agriculture.

Keywords: Wheat, grain, storage proteins, nitrogen, sulfur, nuclear protein, -omics data, biological network

Université Blaise Pascal — Université d'Auvergne

Ecole Doctorale Sciences de la Vie, Santé, Agronomie, Environnement

Thèse de doctorat

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

DOCTEUR D'UNIVERSITÉ
Spécialité : Physiologie et génétique moléculaires

Réponse du grain de blé à la nutrition azotée et soufrée : Etude intégrative des mécanismes moléculaires mis en jeu au cours du développement du grain par des analyses -omiques

---

Par **Titouan Bonnot**

Soutenue le 9 décembre 2016 devant un jury composé de :

TATOUT Christophe	Professeur Université Blaise Pascal, Clermont-Ferrand	**Président**
BUITINK Julia	Directeur de Recherches INRA, Angers	**Rapporteur**
RAJJOU Loïc	Maître de Conférences AgroParisTech, Paris	**Rapporteur**
GALLARDO Karine	Directeur de Recherches INRA, Dijon	**Examinateur**
ZIVY Michel	Directeur de Recherches CNRS, Gif-sur-Yvette	**Examinateur**
BANCEL Emmanuelle	Ingénieur d'Etudes INRA, Clermont-Ferrand	**Co-encadrante**
MARTRE Pierre	Directeur de Recherches INRA, Montpellier	**Co-encadrant**
RAVEL Catherine	Ingénieur de Recherches INRA, Clermont-Ferrand	**Directrice de thèse**

REMERCIEMENTS

Avant de débuter ce manuscrit, je tiens à remercier sincèrement toutes les personnes qui ont contribué de près ou d'un peu plus loin à l'aboutissement de ce projet de thèse.

Tout d'abord je remercie les membres du jury pour avoir accepté de lire et juger mes travaux. Merci à Julia Buitink, Karine Gallardo, Loïc Rajjou, Michel Zivy et Christophe Tatout (et bien sûr Pierre, Catherine et Emmanuelle, également membres du jury, que je remercierai plus bas).

Merci aux membres de mes deux comités de thèse, qui m'ont donné de précieux conseils pour avancer ce projet. Je remercie donc à nouveau Karine Gallardo qui a participé à ces deux comités, puis Colette Larré, Ludovic Bonhomme et Christophe Chambon.

A présent je souhaite adresser un grand merci à toutes les personnes avec qui j'ai eu l'occasion de travailler, en commençant par celles et ceux qui ont été à l'origine de ce projet et qui ont participé à mon encadrement. Et priorité aux personnes qui se sont succédées à la direction de ma thèse. Si l'on suit l'ordre chronologique, merci tout d'abord à Gérard Branlard. J'ai été très honoré d'avoir été l'un de ses derniers thésards. Sa vision de la recherche et ses innombrables connaissances sont pour moi un exemple à suivre. Merci à Pierre Martre, de qui j'ai beaucoup appris. Ta rigueur scientifique et ta capacité d'étendre à l'infini les interprétations des résultats biologiques sont pour moi encore aujourd'hui un mystère et en même temps une grande source d'inspiration. Merci ensuite à Catherine Ravel, qui a fait preuve d'une très grande capacité d'adaptation suite au départ de Pierre et qui a très bien repris la direction de ce projet. Toujours de très bon conseil, je retiendrai ton énergie débordante, ta grande honnêteté et ta capacité de prendre rapidement les bonnes décisions. A présent j'adresse un très grand merci à Emmanuelle Bancel, qui a été au plus près de mon encadrement. Tu es sans aucun doute celle qui a été la plus compréhensive et à l'écoute. Merci pour ta très grande disponibilité, tes nombreux conseils et pour m'avoir toujours proposé ton aide. Et bien sûr merci de m'avoir formé dans le domaine de la protéomique. Pour terminer avec les personnes qui m'ont encadré, merci à Julie Boudet. J'ai beaucoup apprécié tes conseils (je retiendrai ton aide précieuse dans ma préparation pour le concours de l'école doctorale), pour avoir toujours su trouver un moment pour suivre ce projet et m'aider dans son avancement, et pour ton accompagnement lorsque j'ai voulu plonger dans le bassin de l'enseignement.

Je tiens à adresser un très grand merci à l'ensemble des personnes de l'équipe BIG, qui ont participé au climat chaleureux dans lequel s'est déroulée cette thèse. J'ai beaucoup apprécié la période de culture de *Triticum monococcum* qui a mobilisé tout le monde et qui pour moi a été le symbole de l'unité de cette équipe (c'est peut-être un mauvais souvenir pour vous ! Je pense notamment aux notations de floraison le samedi et dimanche matin, ou encore le planning de prélèvement des grains qui était un véritable casse-tête pour toi Annie et toi David ! Un peu speed mais ça fait des bons souvenirs). Donc merci à ceux qui n'ont pas été cités plus haut : Annie, Marielle, Mireille, Sibille, Isabelle et David. David, merci aussi pour les sérieuses parties de fléchettes (si, à ce niveau c'est sérieux !) et les discussions rugby pour déconnecter le temps de midi. Je n'oublie pas bien sûr les filles qui ont été de passage : Ivonne, Céline et Selver, merci pour votre bonne humeur quotidienne.

Merci aux autres personnes, du GDEC et d'ailleurs qui m'ont aidé dans mon travail, et particulièrement Marie Pailloux pour sa disponibilité et sa pédagogie pour expliquer simplement le fonctionnement de RulNet à un novice comme moi ; Christophe Chambon, Didier Viala, Marlène Davanture, Thierry Balliau, Olivier Langella, Michel Zivy pour leur aide et leur expertise en protéomique, Philippe Leroy pour l'annotation fonctionnelle des gènes, sans oublier Richard, Stéphane, Christian, Raphaël et Noël pour la gestion des plantes en serre, et puis Karine pour la gestion des déplacements et Valérie pour la gestion du personnel non permanent. Merci à toutes les autres personnes du GDEC et du site de

Crouël en général, avec qui j'ai eu l'occasion de passer un moment, à la cantine (un grand merci au personnel de la cantine en passant !), aux fléchettes, au ping pong…

Merci à toi Victor pour ton aide sur la fin de ma thèse. Tu as été le bon stagiaire pour aider un thésard de troisième année : autonome et efficace ! Impec !

Merci aux thésards qui ont partagé le même bateau que moi et qui ont souvent été là pendant ces trois ans. Honneur à miss Delphine qui a soutenu en premier, merci pour ta bonne humeur à toute épreuve, ton soutien et les échanges qu'on a pu avoir (de vannes, de ping pong, de volley, de stress… !). Merci Julio, on aura bien rigolé pendant tes séjours à Clermont (je sais que tu es souvent revenu juste pour ça), même si la défaite contre le Portugal est encore dure à digérer (je sais tu y es pour rien mais bon je garde en mémoire ton appel de Lisbonne 5 min après la fin du match…). Merci Robin pour avoir supporté mes vannes (pas drôles ?), pour ton écoute, les moments à la salle (ça a pas duré longtemps ;)) à la machine à caf ou autre (cette fois je crois qu'on va définitivement bosser à deux endroits différents, fallait bien que ça arrive depuis le stage de M1!). Florent merci à toi sur tous les fronts : dans le boulot, aux pauses café, dans les moments de détente en dehors… (et je ne te remercie pas pour le nombre incalculable de roustes que j'ai pu prendre à cause de toi au ping). Merci Bianca (oui je t'intègre dans le paragraphe thésards parce que je fais ce que je veux et je ne peux pas te dissocier de Flo) pour avoir ramené un peu le soleil d'Italie ! Merci à toi Seb, tu as été tellement important pour les moments de déconnexion et de craquage à la salle, au ping-squash, à Marcombes ou autre ! Merci à toi Romain pour les mêmes raisons, le seul regret c'est que l'on n'ait pas débuté la thèse ensemble… Merci Emilie, tu m'auras bien eu en ne voulant jamais parler de la thèse et en m'annonçant deux semaines avant que tu allais rendre le manuscrit ! Bien joué ! Bien joué à toi aussi Auré, tu as bien géré sur la fin, entre Clémence qui s'est mise à la peinture et au tricot à Limoges (dur pour toi d'aller là-bas le week end…) et l'écriture d'un manuscrit de thèse de Physique ! Machine… Je plaisante Clem, merci à toi aussi !

Merci à tous les autres : Thomas, GC et son fillot Antho, Floriane, Estelle, Gaétan et Agathe, Emma et Kiki, Boubou, Yann le PDG, Vélo, Boutet, Jeff. Merci les amis pour les bons moments passés ensemble, à Clermont, Orange, Sauternes, au FreeMusic…

Merci à toute la famille Gillard, avec qui j'ai passé de super week ends et de super vacances. Merci à Christine, Malo, Tristan, Maël, Sylvie, Eric, Nadine, Marion, Tanguy et Romain.

Merci à mes parents. Sans votre soutien et votre compréhension je crois que cela n'aurait jamais été possible, j'en suis sûr même. Vous m'avez toujours fait confiance et encouragé sans jamais me mettre la pression. Merci pour tout. Merci à ma sœur Charlotte. Merci d'avoir expérimenté avant moi les 100 meilleures façons d'être privé de sortie. Je plaisante (à moitié), merci à toi Cha, la vraie grande sœur protectrice toujours là pour moi. C'est bien sûr aussi grâce à toi si j'en suis là. Seb je ne t'oublie pas merci beaucoup à toi, avec Cha vous avez toujours été là, merci pour les concerts, festivals, soirées, séances travaux… Des moments tellement importants pour se défouler. Merci à ma grand-mère Christiane et mon oncle Régis. Merci pour votre accueil sur la région parisienne lors de mes séjours à Gif, et pour les très bons week ends passés ensemble là-bas. A ma famille, un immense merci pour votre soutien.

40,000. Ce chiffre te parlera sûrement puisque c'est toi qui l'as estimé. Pour moi ce manuscrit représente une étape importante sur le plan professionnel, mais marque surtout la fin d'une difficulté sûrement plus grande que la thèse en elle-même. Merci à toi Morgane d'avoir été là. Promis, les prochains 40,000 on les parcourra ensemble… « Crack the code, ….»

Mes derniers remerciements vont vers toi Yann, merci pour tout…

VALORISATION DES TRAVAUX DE THÈSE

Publications

Bancel, E.*, Bonnot, T.*, Davanture, M., Branlard, G., Zivy, M., & Martre, P. (2015). Proteomic approach to identify nuclear proteins in wheat grain. *Journal of Proteome Research*, *14*, 4432–4439. *Co-premiers auteurs

Bonnot, T., Bancel, E., Chambon, C., Boudet, J., Branlard, G., & Martre, P. (2015). Changes in the nuclear proteome of developing wheat (*Triticum aestivum*) grain. *Frontiers in Plant Science*, *6*.

Bonnot, T., Bancel, E., Alvarez, D., Davanture, M., Boudet, J., Pailloux, M., Zivy, M., Ravel, C. and Martre P. Grain subproteome responses to nitrogen and sulfur supply in diploid wheat *Triticum monococcum* ssp. *monococcum*. *Plant Journal.* **(Soumis)**

Bonnot, T., Hatte, V., Bancel, E., Dardevet, M., Leroy, P., Boudet, J., Moing, A., Gibbon, Y., Pailloux, M., Martre, P. and Ravel, C. Integrative response of the einkorn (*Triticum monococcum*) grain to nitrogen and sulfur nutrition. **(En préparation)**

Participation à des congrès

Bonnot, T., Bancel, E., Boudet, J., Chambon, C., Branlard, G and Martre, P. (2013) Analyse des protéines nucléaires du grain de blé tendre. Communication orale au **4ème colloque national Graines,** *Dijon, France*.

Bonnot, T., Bancel, E., Boudet, J., Chambon, C., Branlard, G. and Martre, P. (2015) Changes in nuclear proteome of developing wheat grain. Communication orale au **IV International Conference on Analytical Proteomics**, *Lisbonne, Portugal*.

Branlard, G., Bancel, E., Bonnot, T., Boudet, J., Dardevet, M., Merlino, M., Nadaud, I., Ravel, C. (2015) Fifteen years of proteomics of developmental wheat grain. Communication orale au **IV International Conference on Analytical Proteomics**, *Lisbonne, Portugal*.

Bonnot, T., Bancel E., Alvarez, D., Dardevet, M., Boudet, J., Davanture, M., Moing, A., Gibon, Y., Zivy, M., Pailloux, M., Martre, P. and Ravel, C. (2016) Integration of "omics" data provides tools for understanding effects of variable nitrogen and sulfur supply during wheat grain development**.** Pésentation sous forme de poster au **COST Action FA1306- Diving into integrative cell phenotyping through "omics"**, *Versailles, France*.

LISTE DES ABBRÉVIATIONS ET SIGLES

°Cj	Degrés jours
2D gel	Two-dimensionnal gel electrophoresis
ACN	Acétonitrile
ADN	Acide désoxyribonucléique
ADP	Adénosine diphosphate
alg	Albumines-globulines
ANOVA	Analysis of variance
APS	Adenosine 5'-phosphosulfate
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messager
Asn	Asparagine
Asp	Aspartate
ATP	Adénosine triphosphate
BIG	Biologie Intégrative de la composition du Grain
BLAST	Basic Local Alignment Search Tool
BSA	Bovin serum albumin
CBB	Coomassie Brilliant Blue
CPM	Count Per Million
CS	Cystéine Synthase
DM	Dry Mass
DNA	Deoxyribonucleic acid
DTT	Dithiothréitol
ECL	Enhanced chemiluminescence
ESI	Electrospray Ionization
FACT	Facilitate chromatin transcription
FAO	Food and Agriculture Organization of the United Nations
FM	Fresh Mass
FT	Facteur de Transcription
GAPDH	Glyceraldehyde-3-phosphate deshydrogenase
GDC	Glycine decarboxylase complex
GDEC	Génétique Diversité et Ecophysiologie des Céréales
GLM	GCN4-Like Motif
Gln	Glutamine
Glu	Glutamate
GO	Gene Ontology
GOGAT	Glutamine OxoGlutarate AminoTransferase
GPD	Grain Protein Deviation
GS	Glutamine Synthétase
GSH	Glutathion réduit
GSP	Grain Storage Protein
H_2S	Sulfure d'hydrogène
HATS	High-Affinity Transport System
HDAC	Histone désacétylase
HMG	High-Mobility Group
HMW-GS	High-Molecular-Weight Glutenin Subunit

HPLC	High Performance Liquid Chromatography
HRP	Horseradish peroxidase
HSD	Honestly Significant Different
HSP	Heat Shock Protein
Ile	Isoleucine
IPG	Immobilized pH gradient
IT	Ion Trap
LATS	Low-Affinity Transport System
LC	Liquid Chromatography
LMW-GS	Low-Molecular-Weight Glutenin Subunit
LTQ	Linear Trap Quadrupole
Lys	Lysine
Ma	Million d'années
Mal	Malate
MBD	Methyl-CpG DNA Binding Domain
Met	Methionine
MS/MS	Tandem Mass Spectrometry
Mt	Million de tonnes
N	Azote
N_2	Diazote
NAD	Nicotinamide adénine dinucléotide
NH_4^+	Ammonium
NLS	Nuclear Localization Signal
NO_2^-	Nitrite
NO_3^-	Nitrate
NPC	Nuclear Pore Complex
PB	Prolamine Box
PBF	Prolamine-box Binding Factor
PCA	Principal Component Analysis
PHD	Plant Homeodomain
PMG	Poids de Mille Grains
PMSF	Fluorure de phénylméthylsulfonyle
PR	Protéines de Réserve
q/ha	Quintaux par hectare
RNA	Ribonucleic acid
RNA-Seq	Ribonucleic acid Sequencing
S	Soufre
SDS	Sodium dodécylsulfate
SDS-PAGE	Sodium dodécylsulfate polyacrylamide gel electrophoresis
SO_2	Dioxyde de soufre
SO_4^{2-}	Sulfate
TFA	Trifluoroacetic acid
Trp	Tryptophane
UGP	UDP-Glucose pyrophosphorylase
USDA	United States Department of Agriculture
XIC	Extracted Ion Chromatogram
ZALS	Zwitterionic Acid Labile Surfactant

SOMMAIRE

RÉSUMÉ DES FIGURES

INTRODUCTION GÉNÉRALE

Première source de calories et de protéines dans l'alimentation humaine, le blé est une plante d'intérêt agronomique indéniable, au centre d'enjeux économiques importants. Pour répondre à une demande toujours croissante, l'augmentation des rendements en blé est depuis plusieurs décennies l'objectif majeur des programmes de sélection variétale. Depuis le milieu du XX^{ème} siècle, des progrès conséquents ont été réalisés tant d'un point de vue génétique, que d'un point de vue agronomique avec des changements importants au niveau de la conduite culturale, le développement de la mécanisation, l'utilisation importante de fertilisants notamment azotés et l'emploi de produits phytosanitaires. En France par exemple, ces progrès ont conduit à l'intensification du mode de production responsable d'une remarquable augmentation du rendement en blé tendre qui est passé d'environ 20 quintaux/hectare (q/ha) en 1950 à environ 70 q/ha pendant les années 1990. Cependant, l'utilisation excessive d'engrais azotés, entraînant sa non-assimilation par la plante et son rejet dans l'atmosphère, les rivières et les océans, est à l'origine de nombreuses perturbations aujourd'hui observées au sein des écosystèmes avoisinant les cultures. Face à cette problématique, des mesures ont été prises, par exemple à l'échelle européenne avec la Directive Nitrates qui a pour but de raisonner les apports d'azote et limiter leur impact sur l'environnement. De telles mesures peuvent expliquer l'évolution nettement plus lente des rendements observée depuis quelques années, ce malgré les nombreux progrès génétiques encore réalisés. A cela s'ajoutent la réduction de la diversité des cultures au sein des rotations, avec notamment une réduction de la part des légumineuses avant la culture de blé, les effets du changement climatique, ainsi que la réduction de l'emploi de produits phytosanitaires avec par exemple en France la mise en place du plan Ecophyto. Ces raisons seraient en grande partie responsables du phénomène de stagnation des rendements.

Ces difficultés d'ordre environnemental s'accompagnent d'une forte relation négative existant entre le rendement et la teneur en protéines du grain. Or, la teneur en protéines du grain est un critère important à la base de sa qualité nutritionnelle et technologique. Le maintien d'une quantité de protéines optimale sans impacter le rendement constitue donc un enjeu essentiel, notamment pour la France qui est le 5ème producteur mondial. Outre la teneur, la composition du grain en protéines notamment en protéines de réserve (molécules riches en azote et en soufre) qui sont responsables des propriétés rhéologiques de la farine, influence fortement la qualité d'utilisation du grain. On pourrait ainsi espérer compenser une diminution de la teneur en protéines induite par l'augmentation des rendements et l'évolution des pratiques culturales vers des méthodes plus respectueuses de l'environnement par une composition en protéines de réserve adaptée aux différents usages. Autrement dit, il convient de trouver les solutions pour produire « plus » et « mieux ».

Dans ce contexte, les travaux de l'équipe Biologie Intégrative de la composition du Grain (BIG) ont pour but, entre autres, d'étudier les mécanismes moléculaires impliqués dans le déterminisme de la composition du grain de blé en protéines de réserve et d'identifier des gènes ou régions chromosomiques associés à ces phénomènes de régulation. C'est dans cet axe de recherche que s'est intégré mon projet de thèse. Il avait pour but d'apporter de nouveaux éléments à la compréhension des processus de régulation du développement du grain et de l'accumulation des protéines de réserves en réponse à la nutrition azotée et soufrée. Pour cela j'ai produit ou participé à l'acquisition de différents types de données –omiques sur des grains

en cours de développement produits dans différentes conditions de culture. J'ai en particulier étudié le protéome nucléaire, la synthèse des protéines de réserve étant soumise à une forte régulation transcriptionnelle. Puis en analysant les différents niveaux de données –omiques produites, j'ai étudié la réponse intégrative du grain à la nutrition azotée et soufrée. J'ai débuté ce projet lors de mon stage de Master 2, que j'ai ensuite poursuivi au cours de ces trois années de thèse.

Ce manuscrit débute par une synthèse bibliographique qui fait le point sur les connaissances à l'origine de ce projet, dont la stratégie d'étude et les objectifs sont définis en fin de ce premier chapitre. Les chapitres suivants correspondent aux articles scientifiques publiés au cours de la thèse, soumis ou en préparation. Pour faciliter la lecture, chaque article est accompagné d'une introduction et une conclusion en français. Enfin, une conclusion générale de la thèse ainsi que des éléments de perspectives viennent clore ce manuscrit.

CHAPITRE I :
Synthèse bibliographique

1. LE BLÉ : GÉNERALITÉS

1.1 Origine des blés cultivés

Les blés aujourd'hui cultivés appartiennent au genre *Triticum* et possèdent différents niveaux de ploïdie, qui témoignent des événements de polyploïdisation survenus au cours de leur l'histoire évolutive, avant domestication. On distingue ainsi des espèces diploïdes telles que l'engrain ou petit épeautre (*Triticum monococcum*), des espèces tétraploïdes comme l'amidonnier (*T. turgidum*) et des espèces hexaploïdes (*T. aestivum*, Figure 1a). Les blés tétraploïdes se sont différenciés en plusieurs sous-espèces comme l'amidonnier domestiqué ou le blé dur (*T. turgidum* spp *dicoccoïdes* ou *durum*, respectivement). Il en est de même pour les blés hexaploïdes où l'on distingue deux sous-espèces cultivées, le blé tendre (*T. aestivum* spp *aestivum*) et le grand épeautre (*T. aestivum* spp *spelta*). Le génome des blés hexaploïdes est ainsi constitué de 3 génomes différents : A, B et D, chacun constitué de 7 paires de chromosomes, soit un total de 42 chromosomes (Figure 1b).

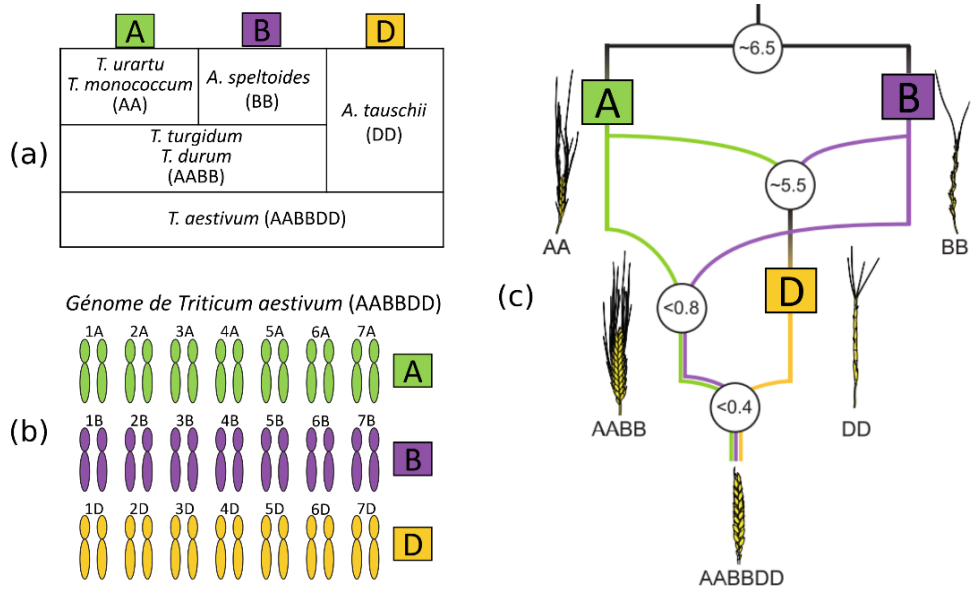


Figure 1. Schéma de l'origine des génomes portés par les différents blés. (a) Catégorisation de différentes espèces de blés en fonction de leur niveau de ploïdie, portant les génomes A, B et/ou D. (b) Schéma du caryotype du blé tendre (*Triticum aestivum* spp *aestivum*) hexaploïde. (c) Schéma de l'histoire évolutive des blés, les dates estimées des événements étant exprimées en millions d'années, d'après Marcussen *et al.* (2014).

La comparaison de ces 3 génomes a permis d'estimer leur date de divergence et de reconstruire l'histoire phylogénétique du blé (Marcussen *et al.*, 2014). Les génomes diploïdes (2n=14) A et B, portés respectivement par *T. urartu* et *Aegylops speltoides* ont divergé il y environ 6,5 Ma puis un premier événement d'hybridation a été à l'origine du génome D (2n=14) il y a 5,5 Ma (Figure 1c). Le génome hexaploïde du blé tendre est alors apparu suite à deux événements majeurs de polyploïdisation. Le premier correspond à une hybridation il y a environ 0,8 Ma entre *T. urartu* (AA) et une espèce proche d'*Ae. speltoides* (BB) qui a permis l'apparition du blé dur sauvage (*T. turgidum*) au génome AABB à l'origine du blé dur actuel. Le second événement a eu lieu il y a environ 0,4 Ma et correspond à un croisement entre *T.*

turgidum (AABB) et le diploïde *Ae. tauschii* (DD), qui a donné naissance à l'ancêtre hexaploïde du blé tendre (AABBDD).

Les premières cultures de blé sont apparues il y a environ 10 000 ans en Mésopotamie au moment de la révolution néolithique (Shewry, 2009). Pour l'Homme, cette période a marqué la transition entre un mode de vie basé sur la chasse et la cueillette à un mode de vie basé sur l'agriculture et l'élevage, entraînant ainsi sa sédentarisation. Sur la base d'éléments archéologiques, génétiques et botaniques, l'origine des premiers blés cultivés que sont l'engrain diploïde (*T. monococcum*) et l'amidonnier tétraploïde (*T. turgidum* spp *dicoccum*) a été estimée à une zone limitée du croissant fertile située à l'amont du Tigre et de l'Euphrate et qui correspond à des territoires de l'actuelle Syrie et de la Turquie (Figure 2, Heun, 1997 ; Lev-Yadun *et al.*, 2000 ; Dubcovsky and Dvorak, 2007). L'origine géographique du blé tendre semble quant à elle être localisée au Nord-Ouest de l'Iran et/ou au Nord-Est de la Turquie.

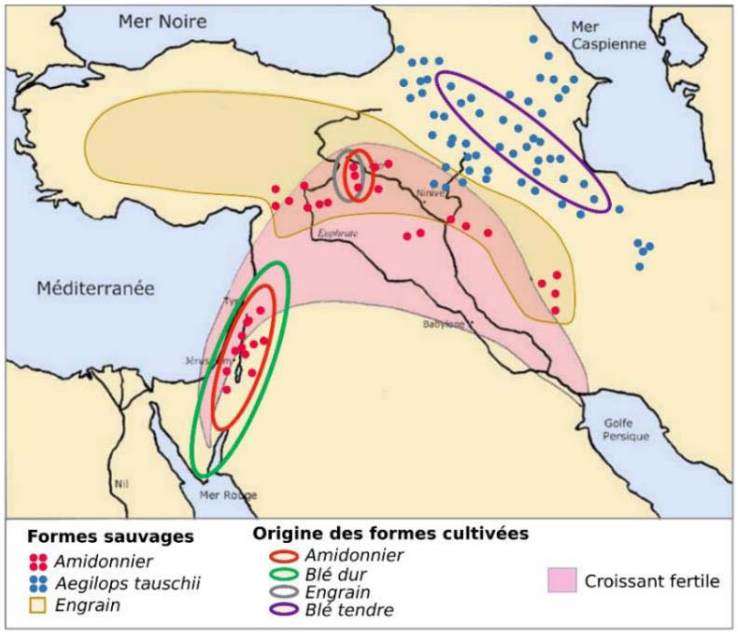


Figure 2. L'origine géographique des blés cultivés. La distribution approximative des formes sauvages d'amidonnier (*T. turgidum* spp *dicoccoïdes*) et d'*Aegilops tauschii* est représentée par des points et celle d'engrain est représentée par la zone jaune. Les zones cerclées correspondent aux régions putatives d'origine des formes cultivées d'amidonnier, de blé dur, d'engrain et de blé tendre. Adapté de Dubcovsky and Dvorak (2007).

La diffusion du blé en Europe a débuté vers environ -8000 ans avant JC à partir du bassin anatolien (Bonjean, 1994). La culture du blé s'est également étendue à l'Est du croissant fertile pour atteindre la Chine vers environ -5000, et à l'Afrique par l'Egypte vers -6000. Plus récemment, le blé fut introduit au XVIe siècle en Amérique par les Espagnols, puis en Australie par les Anglais au XVIIIe siècle, à partir de pools génétiques européens (Shewry, 2009).

1.2 Utilisation des blés

Le blé tendre (*T. aestivum*) est l'espèce la plus cultivée et représente à elle seule 95% de la production mondiale de blé. La domination de cette espèce au sein des cultures s'explique par son fort potentiel de rendement et sa grande plasticité génomique et phénotypique, qui lui permettent d'être cultivée dans la plupart des régions agricoles mondiales (Shewry, 2009). L'utilisation majoritaire du blé tendre provient de son grain, dont la farine panifiable sert à la confection du pain, aliment qui a été à la base de nombreux régimes. Une étude britannique montre que si aujourd'hui les produits céréaliers représentent environ 31% et 23% des apports

journaliers d'énergie et de protéines, respectivement, le pain à lui seul apporte environ 11% des apports pour ces deux catégories (Bates *et al.*, 2011; Shewry and Hey, 2015). La paille qui correspond à la tige coupée pendant la moisson fait également l'objet de diverses utilisations. Elle peut être enfouie après la récolte dans le but de restituer au sol le carbone et les éléments minéraux prélevés par la culture, ou si elle est récoltée, être utilisée comme litière pour les animaux ou servir de fourrage, de combustible, ou d'élément de base pour la fabrication de fumier utilisé comme fertilisant. La seconde espèce de blé la plus produite est le blé dur (*T. durum*), plus adapté au climat sec Méditerranéen et utilisé principalement dans la fabrication de pâtes alimentaires et de semoules. D'autres espèces de blés sont cultivées dans des zones géographiques et sur des surfaces plus restreintes. C'est le cas de l'engrain (*T. monococcum*), l'amidonnier (*T. turgidum* spp *dicoccoïdes*) et le grand épeautre (*T. spelta*). Les cultures d'engrain sont limitées à des zones essentiellement à proximité de la région méditerranéenne (Turquie, pays des Balkans, sud de l'Italie et de la France, Espagne, Maroc). Cette espèce diploïde est très appréciée pour ses excellentes propriétés nutritionnelles et ses forts taux de protéines et de caroténoïdes (Corbellini *et al.*, 1999 ; Hidalgo *et al.*, 2006 ; Hidalgo and Brandolini, 2014).

1.3 La Production

Aujourd'hui le blé est la seconde céréale la plus produite dans le monde et représente 33% de la production céréalière totale, derrière le maïs (46%) et devant le riz blanchi (21%, USDA, 2016). Au cours de la campagne 2014/2015, une production record a été réalisée avec 730 millions de tonnes (Mt) produites (FAO, 2016), entraînant une augmentation du stock mondial estimé aujourd'hui à 200 Mt. La France représente environ 6,2% de la production mondiale de blé et se situe à la première place des producteurs au niveau européen et à la 5ème place au niveau mondial, les premiers étant la Chine, l'Inde, les Etats-Unis et la Russie (Figure 3).

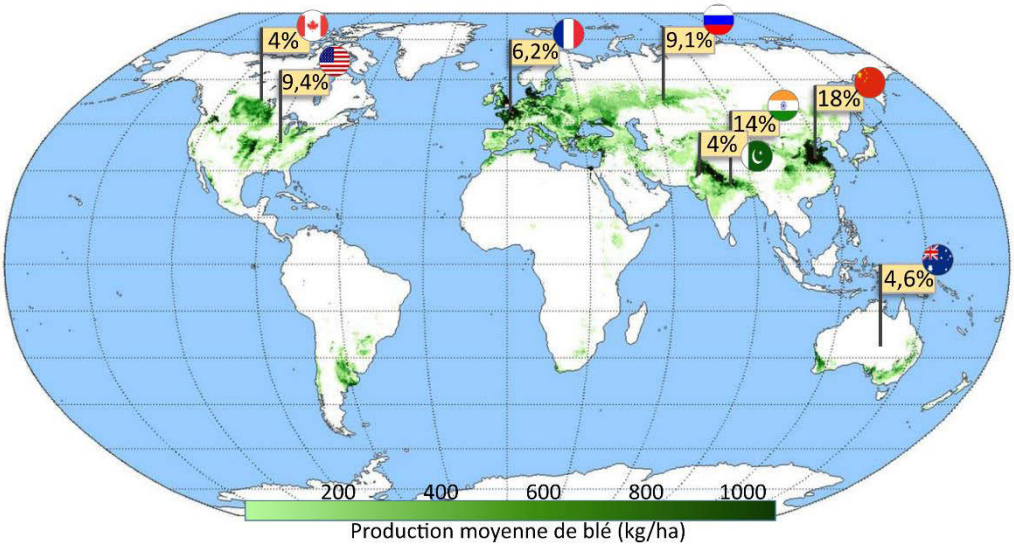


Figure 3. Principales zones et pays de production de blé dans le monde. D'après (Monfreda *et al.*, 2008). La part des 8 principaux pays exportateurs dans la production mondiale a été indiquée. Source : FAOSTAT (http://faostat.fao.org/site/339/default.aspx).

En France, les surfaces arables allouées à la culture de blé sont de 5,3 Mha, soit 2,3% des surfaces mondiales consacrées à cette culture. Ces chiffres rendent compte des deux types d'agriculture pratiqués dans le monde. En France comme en Europe, l'agriculture privilégiée est de type intensive, synonyme de fort rendement (avoisinant les 60 q/ha) en moyenne en Europe et 75 q/ha en France) sur des surfaces cultivées réduites, qui diffère de l'agriculture extensive pratiquée par exemple aux Etats Unis où les surfaces allouées sont beaucoup plus importantes mais où les rendements sont deux fois moins élevés, avec environ 29 q/ha (FranceAgrimer, 2015).

1.4 Conduite culturale et développement de la plante

Pour atteindre un haut niveau de production, différents traitements et apports à la culture doivent être réalisés pour subvenir aux besoins de la plante tout au long de son cycle de développement et ainsi obtenir le rendement et la qualité du grain espérés. On distingue couramment deux grandes phases dans le développement d'une plante de blé : la phase végétative au cours de laquelle il y a production de tiges, feuilles et racines et la phase reproductive qui correspond à l'apparition d'un épi portant des fleurs puis à la formation des grains après fécondation. La transition entre ces deux phases diffère entre les blés dits « d'hiver » qui sont semés à l'automne et qui nécessitent une période de froid, appelée vernalisation, pour initier la formation de leur appareil reproducteur, et les blés dits « de printemps » qui ne nécessitent pas cette période de vernalisation (Acevedo *et al.*, 2002).

Après germination et levée de la graine, l'étape de tallage consiste en l'apparition de tiges secondaires appelés talles (Figure 4). Pendant la phase de montaison, il y a une croissance rapide des entre-nœuds. L'épi monte progressivement dans la gaine, provoquant son gonflement, puis c'est l'épiaison généralement au mois de juin. Une fois la formation des organes floraux achevée, l'étape de floraison marque le début du développement du grain, qui est ensuite récolté au mois d'août lorsqu'il a atteint la maturité.

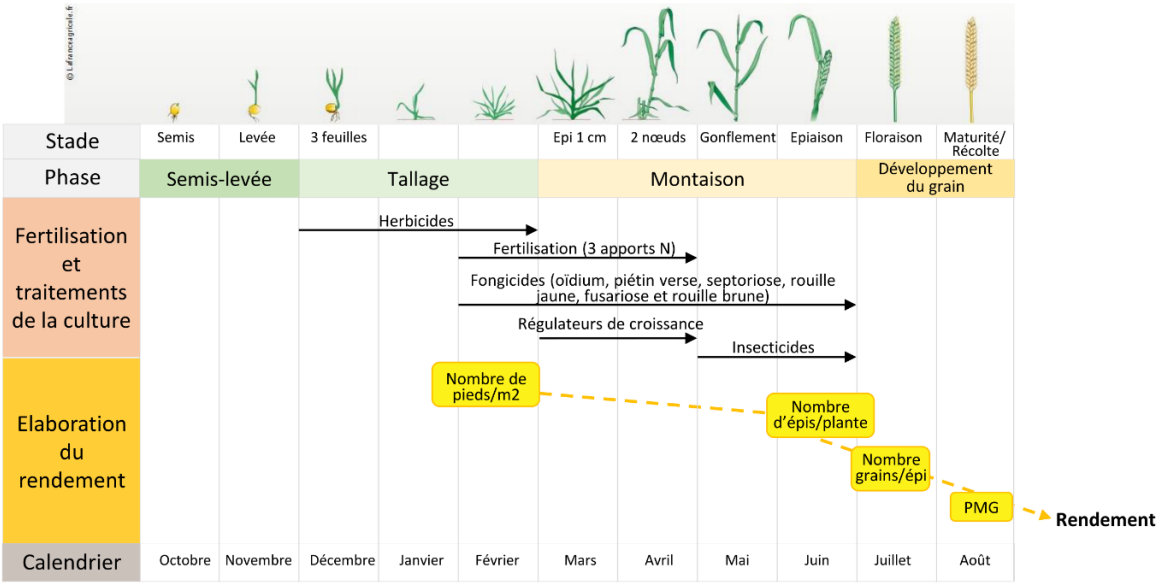


Figure 4. Schéma simplifié du développement d'une plante de blé tendre d'hiver et de la mise en place du rendement en grains (Arvalis-Institut du végétal, 2016).

Les étapes clés de tallage et de montaison sont très sensibles à l'environnement et, dans le cas d'une agriculture conventionnelle, s'accompagnent de plusieurs traitements phytosanitaires pour lutter contre les adventices, maladies fongiques et autres ravageurs nuisibles à la culture (Figure 4). Afin de subvenir aux besoins azotés de la plante tout au long de son cycle, trois apports sont généralement pratiqués : au stade de tallage, au stade épi 1 cm et enfin, entre le stade 2 nœuds et le stade de gonflement. Un apport tardif après épiaison peut également être pratiqué et favorise une teneur en protéines des grains plus élevée. Le fractionnement des apports est jugé le plus efficace pour obtenir de hauts rendements. En effet, le rendement se met en place tout au long du cycle de production, à travers différentes composantes (Figure 4). A l'issue du tallage, le nombre de pied par m^2 est défini et constitue la première composante. Les autres composantes sont le nombre d'épis par plante, déterminé pendant l'épiaison, le nombre de grains par épi qui dépend du nombre d'épillet par épi, du nombre de fleurs par épillet et du pourcentage de fertilité et le poids de mille grains (PMG).

1.5 Rendement versus quantité de protéines

Chez le blé, les programmes d'amélioration variétale ciblent principalement un fort potentiel de rendement. Malgré ces efforts, on assiste depuis la fin des années 1990 à leur stagnation qui ne peut être expliquée par une absence des progrès génétiques (Graybosch and Peterson, 2010; Brisson *et al.*, 2010). En fait, cette stagnation du rendement serait due au changement climatique d'une part et à la diminution de l'application de fertilisants azotés d'autre part (Brisson *et al.*, 2010). Face à cette problématique, associée à une diminution des terres arables et une croissance constante de la population mondiale, l'augmentation de la production de blé représente un défi de premier plan. Une seconde difficulté demeure dans la forte relation négative existant entre le rendement et la teneur en protéines du grain (Simmonds, 1995 ; Feil, 1997 ; Brancourt-Hulmel *et al.*, 2003 ; Bogard *et al.*, 2010 ; Figure 5).

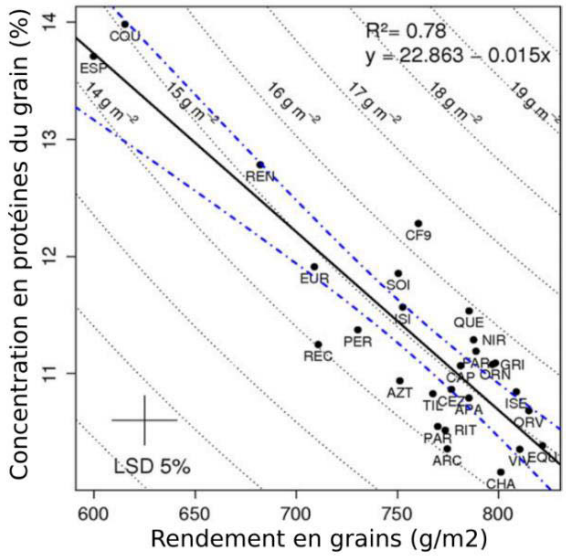


Figure 5. Relation négative entre la concentration en protéines et le rendement en grains. Les données sont les moyennes pour 27 génotypes cultivés dans 27 environnements différents. La droite noire représente la régression linéaire du taux de protéines en fonction des valeurs de rendement. Les lignes en pointillés bleus représentent l'intervalle de confiance à 95% de la régression. Les lignes en pointillés noirs représentent les courbes iso-protéines-rendement en g/m^2. D'après Bogard *et al.* (2010).

Cette relation négative est soumise à de forts effets génétiques et environnementaux ainsi qu'à des interactions entre génotypes et environnements. Certains génotypes présentent un écart

à la relation teneur en protéines-rendement en grains (ou GPD, Grain Protein Deviation ; Figure 5). Autrement dit, le GPD permet d'identifier les génotypes ayant une concentration en protéines supérieure à celle prédite par cette relation (Monaghan *et al.*, 2001). Ceux associant un fort potentiel de rendement et une forte concentration en protéines représentent ainsi des blés très compétitifs sur le marché de l'exportation. Ainsi pour favoriser leur inscription au catalogue des blés, de telles variétés se voient attribuer un bonus par les instances décisionnelles.

Pour l'export, un minimum de 11,5% de protéines par grain est exigé pour entre autres, répondre au mieux aux exigences des industriels de la transformation, le blé étant en effet consommé principalement après transformations. Néanmoins, la quantité de protéines n'est pas le seul critère d'évaluation des grains de blé. En effet, la première transformation que subit le blé est la mouture. Or la valeur meunière repose largement sur la dureté du grain qui détermine l'énergie pour réduire le grain en farine, la granulométrie des farines obtenues ou encore le pourcentage d'amidon endommagé. La qualité technologique du grain est donc appréciée par plusieurs critères (mesure de la dureté, composition, ou humidité). Elle permet de départager les blés dans la forte diversité de l'offre mondiale et représente donc un critère important à prendre en compte dans les programmes d'amélioration variétale.

2. LE GRAIN : STRUCTURE ET COMPOSITION

2.1 Phases du développement du grain

Le développement du grain de blé débute à l'issue de la double fécondation (soit l'anthèse) de l'ovule. Un des deux gamètes mâles fusionne avec les deux noyaux polaires du sac embryonnaire pour former l'albumen (3n) et le second gamète mâle féconde l'oosphère pour former l'embryon (2n). Les stades de développement peuvent être estimés à partir du calcul du nombre de jours après anthèse, ou en calculant le temps thermique exprimé en degré-jours (°Cj) après anthèse, qui correspond à la somme des températures moyennes journalières accumulées par le grain. Ce dernier mode de calcul, qui tient compte des variations de température pouvant survenir au cours du développement du grain a été choisi dans cette thèse pour estimer de manière précise les différents stades. Le développement du grain est classiquement décomposé en trois phases principales : la phase de cellularisation, de remplissage et de maturation (Figure 6, Sabelli and Larkins, 2009).

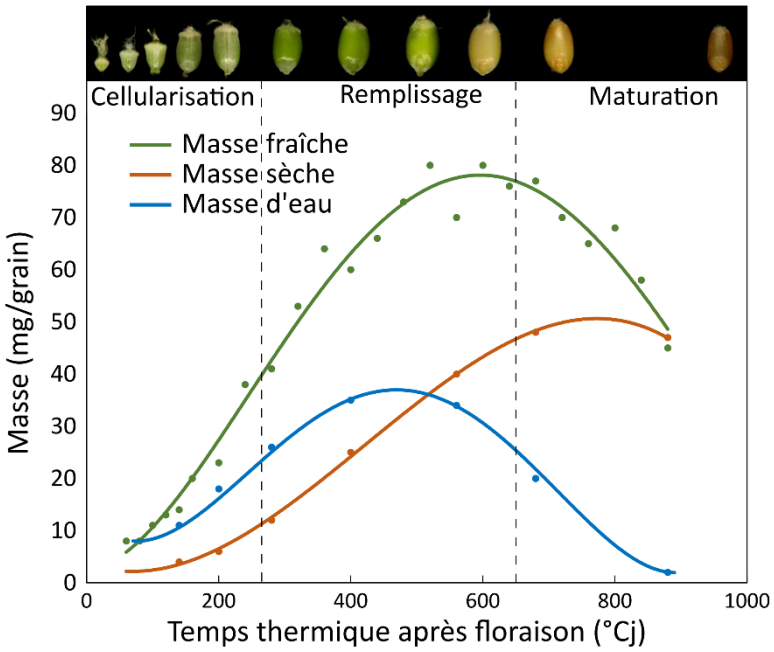


Figure 6. Variations des masses de matière fraiche (vert), matière sèche (orange) et d'eau (bleu) au cours du développement du grain. Le développement du grain est divisé en trois phases : cellularisation, remplissage et maturation.

Après la double fécondation, l'albumen subit des divisions successives sans cytokinèse qui aboutissent à environ 70°Cj à la formation d'un coenocyte à plus de 2000 noyaux répartis à la périphérie de celui-ci, autour d'une large vacuole centrale (Mares *et al.*, 1975). S'ensuit une étape de cellularisation et de divisions cellulaires jusqu'à environ 220-250°Cj (Chojecki *et al.*, 1986), caractérisée par une forte accumulation d'eau (Figure 6). Le grain acquiert alors sa taille finale. Pendant la phase de remplissage qui suit, le grain accumule rapidement ses composés de réserve, essentiellement de l'amidon et des protéines. La vitesse d'accumulation des réserves diminue vers 550°Cj pour devenir nulle généralement vers 650-700°Cj, lorsque la concentration en eau du grain est proche de 45g pour 100g de matière fraîche (Schnyder and Baum, 1992). La dernière phase appelée maturation ou dessiccation correspond à une perte d'eau rapide, entraînant une réduction des activités métaboliques (Young and Gallie, 2000). Le grain contient alors une quantité de masse sèche maximale et atteint l'état dit de maturité physiologique.

2.2 Structure du grain de blé

Le grain de blé est un fruit sec indéhiscent appelé caryopse, constitué de trois parties : l'embryon, l'albumen et les couches périphériques qui représentent environ 3%, 83% et 14% de la masse sèche du grain, respectivement (Figure 7, Pomeranz, 1988; Barron *et al.*, 2007).

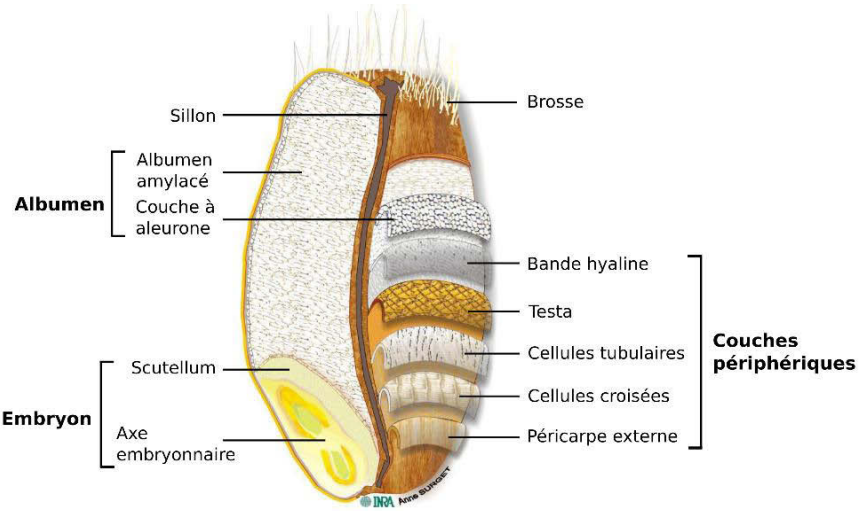


Figure 7. Anatomie du grain de blé tendre. Le grain de blé est constitué de trois parties : l'embryon, l'albumen et les couches périphériques. D'après Surget and Barron (2005).

L'embryon est constitué de l'axe embryonnaire qui donnera la tigelle, le mésocotyle et la radicule de la future plantule ainsi que du scutellum ou cotylédon qui constitue une zone d'échange entre l'embryon et l'albumen (Evers and Millar, 2002). L'albumen correspond au tissu de réserve et est constitué de l'albumen amylacé et de la couche à aleurone. La couche à aleurone est composée généralement d'une couche de cellules et assure un rôle nourricier et protecteur. Après imbibition du grain, les gènes d'enzymes protéolytiques et hydrolytiques sont activés dans ces cellules pour digérer les parois cellulaires de l'albumen et permettre la mobilisation des réserves pour la croissance de l'embryon (Sabelli and Larkins, 2009). Les couches périphériques sont constituées de l'intérieur vers l'extérieur par la bande hyaline, la testa, les cellules tubulaires, les cellules croisées et le péricarpe externe. Les cellules tubulaires et cellules croisées forment le péricarpe interne. Pendant le développement du grain, le péricarpe évite les pertes d'eau sans pour autant empêcher sa pénétration (Evers *et al.*, 1999).

2.3 Composition du grain de blé

A maturité, le grain de blé est essentiellement constitué d'amidon, qui représente 65-70% de la masse sèche totale (Figure 8a). Cet amidon est localisé au sein de l'albumen amylacé, où il s'accumule dans les amyloplastes. Il constitue la principale ressource énergétique de la plantule pendant le processus de germination. Les protéines représentent 10 à 13% de la matière sèche et se retrouvent quant à elles dans tous les tissus, avec une concentration plus importante dans l'embryon et la couche à aleurone. Amidon et protéines s'accumulent pendant la phase de remplissage, selon un profil similaire dans le temps (Figure 8b, Laudencia-Chingcuanco *et al.*, 2007; Shewry *et al.*, 2012). Les pentosanes (polysaccharides non amylacés) sont les principaux constituants des parois cellulaires de l'albumen. Ces fibres représentent environ 2 à 3% de la

masse sèche. Enfin, les lipides, la cellulose, les minéraux et les vitamines sont des constituants mineurs du grain.

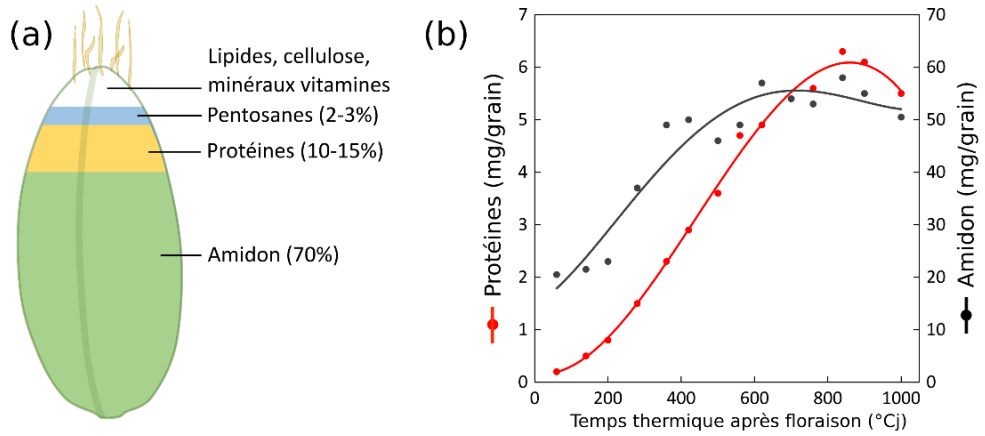


Figure 8. Composition du grain de blé et accumulation des réserves (amidon et protéines) au cours du développement. (a) Représentation de la part des différents composants du grain mature, exprimés en pourcentage de la masse sèche du grain. (b) Cinétique d'accumulation de l'amidon (noir) et des protéines (rouge). D'après Laudencia-Chingcuanco *et al.* (2007).

2.4 Les protéines du grain

L'étude des protéines du grain de blé remonte à 250 ans, lorsque des protéines du gluten ont été isolées pour la première fois en 1745. En 1924, Osborne a proposé une classification des protéines en fonction de leur solubilité : les albumines (solubles dans l'eau), les globulines (solubles dans les solutions salines), les gliadines (solubles dans de l'alcool 70%) et les gluténines (solubles dans une base ou un acide ou des détergents en présence d'un réducteur ; Osborne, 1907; Osborne, 1909). Cette classification a été revue par Shewry et ses collaborateurs (1986) qui ont proposé deux groupes principaux : les protéines métaboliques et les protéines de réserve (PR ; Figure 9).

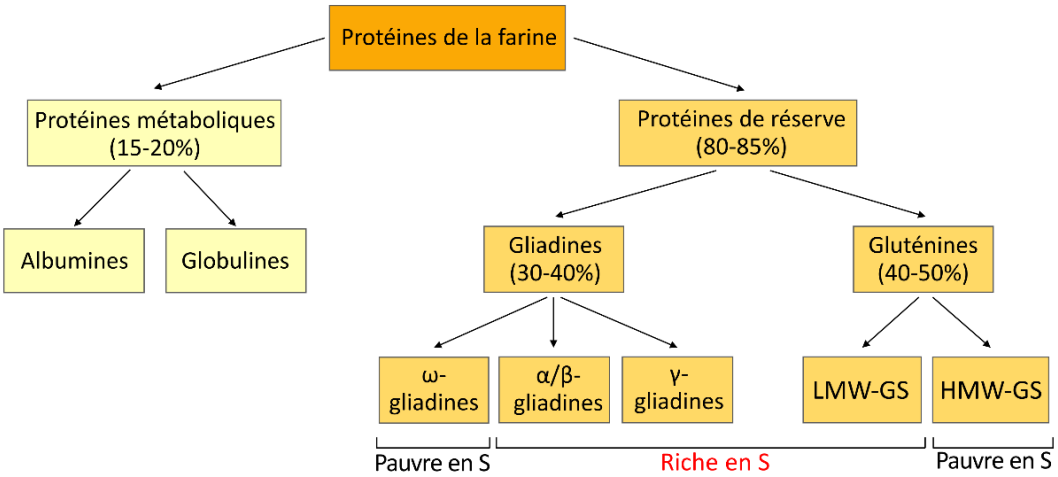


Figure 9. Classification des protéines de la farine du grain de blé : rapprochement entre les classifications d'Osborne et Shewry. La part de chaque classe de protéine est exprimée en pourcentage par rapport aux protéines totales. D'après Feillet (2000).

Les protéines métaboliques correspondent aux albumines et globulines et représentent 15 à 20% des protéines totales. Elles sont constituées d'un grand nombre de protéines ayant des fonctions diverses, telles que des protéases, hydrolases, oxydo-réductases ou des inhibiteurs d'enzymes (Feillet, 2000). Elles participent à la formation du grain et à l'accumulation des réserves dans l'albumen (Vensel *et al.*, 2005).

Les PR, aussi connues sous le nom de prolamines de par leur richesse en proline et glutamine, représentent 80 à 85% des protéines totales de la farine. Elles sont divisées en deux groupes, les gliadines et les gluténines et sont les constituants du gluten, à l'origine des propriétés rhéologiques de la farine. Les gluténines forment des polymères et sont responsables des propriétés viscoélastiques du gluten alors que les gliadines sont des protéines monomériques qui confèrent l'extensibilité de la pâte (Branlard *et al.*, 2001). Les gluténines sont divisées en deux fractions, à savoir les gluténines de faible poids moléculaire (LMW-GS) et de haut poids moléculaire (HMW-GS). Les gliadines sont quant à elles divisées en différentes classes : α/β-, γ- et ω-gliadines.

Les PR sont produites par des loci relativement bien connus, exprimés spécifiquement dans l'albumen (Figure 10). Les gènes codant les HMW-GS sont situés aux loci homéologues *Glu1* sur le bras long des chromosomes du groupe 1 (Payne *et al.*, 1980). Chaque locus est constitué de deux gènes étroitement liés. La synthèse de D'Ovidio and Masci (2004) rappelle que les LMW-GS sont codées par les trois loci homéologues *Glu3* sur le bras court des chromosomes du groupe 1. Des loci supplémentaires ont cependant été détectés. On estime ainsi que le génome du blé compte 30 à 40 gènes codant ces PR. Les gliadines sont produites par des loci complexes situés sur les bras courts des chromosomes des groupes 1 et 6 (Payne *et al.*, 1984). Les α/β- et quelques γ-gliadines sont codés par les loci *Gli-2* des chromosomes 6. Les ω-gliadines et la majorité des γ-gliadines sont produites par des loci des chromosomes 1. Plus précisément, trois loci *Gli-1* codent les γ- et ω-gliadines. Ces dernières sont aussi produites par *Gli-A3* et *Gli-B3*. A côté de ces principaux loci, d'autres plus rares ont été mis en évidence comme *Gli-A5* et *Gli-A6* (Metakovsky *et al.*, 1996). Comme les LMW-GS, les gliadines sont codées par des loci formés de clusters de gènes avec des séquences proches ou identiques. On estime qu'il y a par génome, en incluant les pseudo-gènes, 75 à 450 gènes codant les α/β-gliadines, 45 à 120 gènes codant les γ-gliadines et enfin 45 à 60 codant les ω-gliadines. Les gènes de ces clusters sont transmis en blocs alléliques.

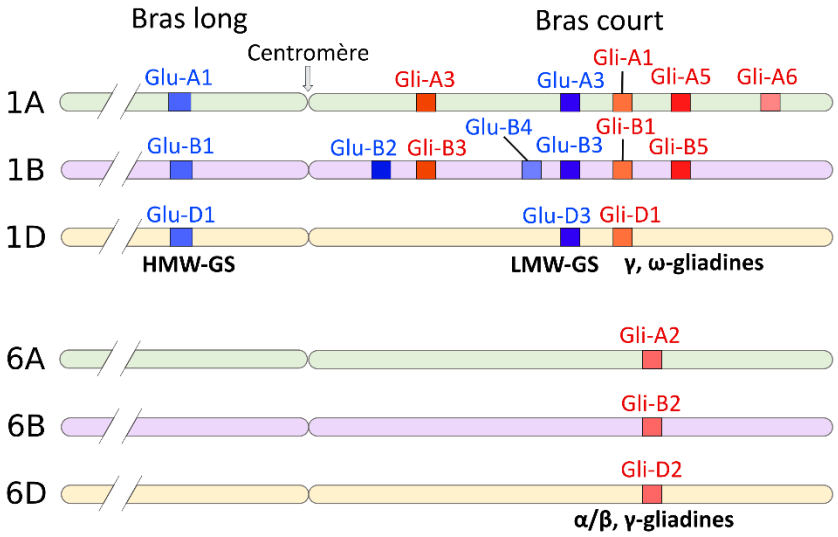


Figure 10. Représentation schématique des principaux loci des gènes de protéines de réserve chez le blé. D'après Shewry and Halford (2003).

2.5 Influence de l'environnement sur le développement et la composition du grain

L'environnement de la plante de blé influence la durée du développement du grain, sa teneur et sa composition en protéines, mais aussi la polymérisation des protéines et le rendement (Nuttall *et al.*, 2016). La température, l'accès à l'eau et la fertilisation sont les facteurs environnementaux qui impactent le plus ces paramètres (Malik, 2009). Les effets de fortes températures, particulièrement pendant le remplissage du grain, sont assez bien caractérisés (Branlard *et al.*, 2015 ; Asseng *et al.*, 2015). Ce type de contrainte entraîne une augmentation des enzymes de la glycolyse et une diminution de l'ADP glucose pyrophosphorylase, l'enzyme qui catalyse la première étape de la voie de biosynthèse de l'amidon. En conséquence, la réduction d'accumulation d'amidon est responsable d'une diminution du rendement. Par ailleurs, de nombreuses protéines chaperonnes essentiellement des HSP, ainsi que des protéines de défense sont fortement produites en réponse à de fortes températures (Skylas *et al.*, 2002; Majoul-Haddad *et al.*, 2013).

Les PR sont quant à elles moins affectées par le stress thermique que les protéines métaboliques (Majoul-Haddad *et al.*, 2013). Toutefois en condition de fortes températures, la proportion de gliadines par rapport aux gluténines est augmentée (Dupont and Altenbach, 2003). Il a également été montré que l'expression des gènes de PR est initiée plus tôt, mais la durée d'accumulation des protéines correspondantes est raccourcie (Altenbach *et al.*, 2002), comme pour l'accumulation de l'amidon. Ces observations sont en réalité dues à un développement du grain accéléré en condition de fortes températures. Des résultats similaires sont observés dans le cas d'un stress hydrique, qui entraîne une taille et un poids de grain réduits. Les effets sur le grain sont accentués en cas de combinaison de ces deux types de stress, hydrique et thermique (Nicolas *et al.*, 1984).

La nutrition de la plante joue également un rôle très important dans le développement du grain et la mise en place de ses réserves. Outre leur catégorisation basée sur les classifications d'Osborne et de Shewry, les PR sont également classées selon leur composition en acides aminés soufrés, cystéine et méthionine. Ainsi on distingue les PR riches en soufre (α/β- et γ-gliadines, LMW-GS) de celles pauvres en soufre (ω-gliadines et HMW-GS, Shewry *et al.*, 1997; Shewry *et al.*, 2001). La synthèse de ces protéines est ainsi fortement influencée à la fois par la disponibilité en azote (N) et en soufre (S) dans le sol.

3. L'AZOTE ET LE SOUFRE : DE LA NUTRITION DE LA PLANTE AUX PROTÉINES DE RÉSERVE DU GRAIN

3.1 Métabolisme de l'azote et du soufre dans la plante

- Transport et assimilation de l'azote

La nutrition azotée des plantes correspond majoritairement à une absorption racinaire d'N minéral sous forme de nitrate (NO_3^-) et d'ammonium (NH_4^+). Dans certains cas de symbioses, la plante, peut également utiliser l'N gazeux (N_2). Parmi ces symbioses, la plus connue est celle observée chez les Fabacées mettant en jeu des bactéries du genre *Rhizobium*. Chez le blé, il a été observé des associations avec des bactéries fixatrices d'N atmosphérique du genre *Frankia*. Les plantes peuvent encore utiliser l'N organique présent dans le sol (acides aminés, courtes chaines peptidiques).

La plante de blé absorbe préférentiellement l'N sous forme de nitrate (Hirel *et al.*, 2007). Cette absorption met en jeu deux types de transporteurs : des transporteurs de haute affinité (Hig-affinity transport system, HATS) et les transporteurs de faible affinité (Low-affinity transport system, LATS). Les LATS appartiennent à la famille des NRT1, qui compte 53 membres connus chez *Arabidopsis thaliana* (Masclaux-Daubresse *et al.*, 2010) et 16 gènes ont récemment été identifiés chez le blé (Buchner and Hawkesford, 2014). Le premier gène caractérisé au sein de cette famille est le gène NRT1.1, qui est à la fois transporteur et senseur de nitrates. Sa mutation inhibe l'élongation des racines de la plante dans les zones de sol riches en nitrates, entraînant une réduction considérable de l'absorption d'N (Remans *et al.*, 2006). Les HATS sont quant à eux des transporteurs de la famille des NRT2 qui compte sept membres chez *A. thaliana*, dont l'expression est inhibée en présence d'une concentration élevée en nitrates dans le sol (Girin *et al.*, 2007). Ce système de transporteurs de nitrates de haute ou faible affinité permet à la plante de modifier rapidement sa capacité d'absorption des nitrates en fonction de ses besoins et de la disponibilité en N dans la solution du sol.

Les nitrates absorbés sont ensuite soit stockés dans la vacuole, soit assimilés dans les racines ou les feuilles après transport dans le xylème. L'assimilation de l'N consiste en deux étapes successives de réduction des nitrates, conduisant à la formation de nitrites (NO_2^-) grâce à l'action de la nitrate réductase dans le cytosol, puis d'ammonium sous l'action de la nitrite réductase dans le chloroplaste (Figure 11).

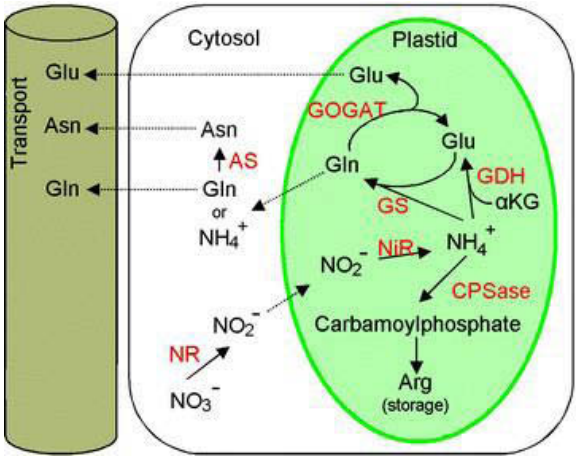


Figure 11. Schéma de l'assimilation de l'azote chez les plantes. NR : Nitrate Réductase ; AS : Asparagine Synthétase ; NiR : Nitrite Réductase ; GS : Glutamine Synthase ; GOGAT : Glutamate Synthase ; GDH : Glutamate Déshydrogénase ; CPSase : Carbamoylphosphate Synthétase. D'après Miura (2013).

L'ammonium est ensuite rapidement incorporé à des acides organiques pour former des acides aminés. Pour cela, la principale voie est le cycle GS/GOGAT, qui correspond à la transformation de l'ammonium en glutamine sous l'action de la glutamine synthétase (GS) puis la formation de glutamate par la glutamate synthase (GOGAT, Miflin and Habash (2002)). Glutamine, glutamate et asparagine sont les principales formes de transport de l'N dans la plante.

- Transport et assimilation du soufre

Le S existe dans le sol sous la forme minérale à l'état de sulfates (SO_4^{2-}) et de dihydrogène de S (H_2S), et sous la forme organique, qui peut représenter jusqu'à 99% du S total disponible. Le sulfate du sol est la forme majoritaire de S absorbée par la plante, bien qu'elle soit également capable d'absorber au niveau foliaire le dioxyde de S (SO_2) provenant des rejets industriels. La quantité de sulfate disponible dépend du reliquat non lessivé en sortie d'hiver, de la minéralisation de la matière organique en début de printemps et de la retombée atmosphérique de dioxyde de S. Le sulfate, absorbé au niveau racinaire par la plante, est transporté jusqu'aux chloroplastes des feuilles où il est assimilé. La famille des transporteurs de sulfate comprend 14 membres chez *A. thaliana* et 10 ont pu être caractérisés chez le blé (Hawkesford, 2003 ; Buchner *et al.*, 2010). Comme l'N, le S peut également être stocké transitoirement dans la vacuole des cellules racinaires et foliaires sous forme de sulfate pour une utilisation ultérieure (Kataoka, 2004).

Pour être assimilé, le sulfate est dans un premier temps activé par l'ATP sulfurylase pour former de l'adénosine 5'-phosphosulfate (APS ; Hirase and Molin, 2003 ; Figure 12). Sous l'action de l'APS réductase, l'APS est convertie en sulfite (SO_3^{2-}) puis en ion sulfure (S_2^-). La biosynthèse de cystéine est ensuite divisée en deux étapes : la sérine acétyltransférase catalyse l'acétylation de la sérine, puis la cystéine synthase (CS) utilise l'*O*-acétylsérine ainsi formé et l'ion sulfure S_2^- pour synthétiser la cystéine. La CS joue donc un rôle majeur puisqu'elle est responsable de la formation de la cystéine, qui tient une place centrale dans la voie d'assimilation du S (Figure 12). En effet, la méthionine et le glutathion (GSH) sont tous deux formés à partir de cet acide aminé.

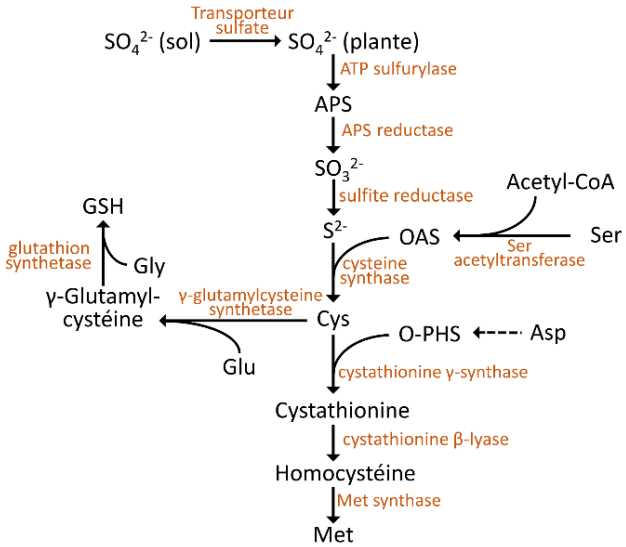


Figure 12. Schéma des voies d'assimilation du soufre dans la plante. APS : Adenosine 5'-Phosphosulfate ; GSH : Glutathion ; OAS : O-acétylsérine ; O-PHS : O-phosphohomosérine. D'après Hirase and Molin (2003).

- Interactions entre l'azote et le soufre

Du fait du rôle central de l'N et du S dans la synthèse des acides aminés et donc des protéines, il existe de fortes interactions entre les assimilations du sulfate et du nitrate dans la plante, mises en évidence depuis de nombreuses années (Figure 13, Reuveny *et al.*, 1980 ; Zhao *et al.*, 1993 ; Kopriva and Rennenberg, 2004 ; Jamal *et al.*, 2010). Une insuffisance de S dans le sol limite l'efficacité de l'apport d'N à la culture (Fazili *et al.*, 2008). Il a également été démontré chez *A. thaliana* qu'une déficience en N affecte le niveau d'expression des gènes impliqués dans l'assimilation du sulfate, suggérant une régulation de cette voie par l'N au niveau transcriptionnel (Yamaguchi *et al.*, 1999 ; Koprivova *et al.*, 2000). Du fait de ces interactions, l'ajout d'N et de S de manière concomitante est donc nécessaire pour maximiser l'assimilation et l'accumulation de N et de S dans la plante et dans le grain.

L'N et le S sont également tous deux impliqués dans des mécanismes de réponse aux stress biotiques et abiotiques (Giordano and Raven, 2014). Ces réponses impliquent par exemple des composés comme le glutathion, dont l'abondance dépend à la fois d'N et de S, et qui joue un rôle central dans la défense de la cellule contre les stress oxydatifs (Noctor *et al.*, 2002). Des interactions existent également avec le métabolisme du carbone (C). En effet, l'absorption des nitrates et du sulfate au niveau racinaire est un processus actif qui nécessite des apports énergétiques issus du métabolisme du carbone (Figure 13). La disponibilité en carbone est donc un facteur de régulation des transporteurs d'N et de S.

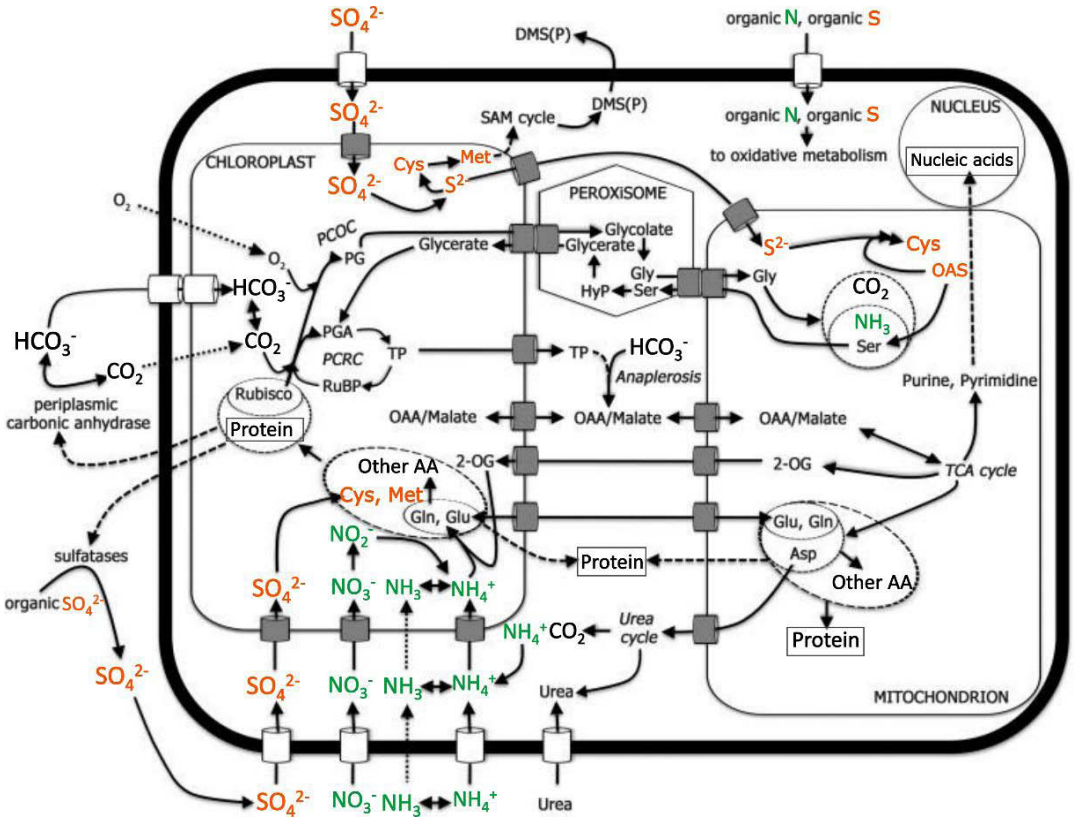


Figure 13. Principales interactions entre les métabolismes du C, N et S chez les plantes. Les flèches pleines indiquent les voies connues, les flèches hachurées indiquent celles qui ne sont pas encore totalement caractérisées et les flèches en pointillés représentent la diffusion du CO_2, O_2 et NH_3. D'après Giordano and Raven (2014).

3.2 Remobilisation de l'azote et du soufre vers le grain

La majeure partie de l'N du grain de blé est issue de la remobilisation de l'N assimilé plus tôt dans les parties végétatives (Austin *et al.*, 1977). Il a été estimé que cette source représente en moyenne 70% de l'N total du grain mature, ce chiffre pouvant varier de 60 à 95% suivant le génotype et les conditions de culture (Palta and Fillery, 1995 ; Kichey *et al.*, 2007). La remobilisation d'N des organes sources vers les organes puits est un mécanisme qui intervient pendant le stade végétatif, des feuilles en sénescence vers les feuilles en développement, ou pendant le stade reproductif, des feuilles sénescentes vers les grains (Malagoli *et al.*, 2005). Il a été montré que ce deuxième type de remobilisation est principalement régulé par la disponibilité en N dans les organes sources (Martre *et al.*, 2003). Les feuilles constituent une source importante car elles sont riches en protéines, notamment la Rubisco qui est présente en abondance dans les plastes. La remobilisation correspond à une mort cellulaire programmée au cours de laquelle ces protéines sont dégradées. Les enzymes intervenant dans le métabolisme des acides aminés permettent alors de convertir les acides aminés issus de cette dégradation des protéines en glutamine et asparagine, qui représentent les deux principales formes azotées transportées vers le grain. La seconde source d'N pour le grain correspond à une absorption et une assimilation directement pendant le remplissage du grain, d'où l'intérêt d'un apport tardif en N. Après floraison de l'épi, il a été montré que la taille et la quantité d'N du grain peuvent être significativement réduits en cas d'une condition de déficience en N (Dupont and Altenbach, 2003).

Dans le grain mature, le S est présent majoritairement sous forme de glutathion, de cystéine et de méthionine. Contrairement à l'N qui est principalement issu de la remobilisation des feuilles vers le grain, le S total du grain de blé est issu à moins de 50% de la remobilisation, les formes principales de S remobilisées étant le glutathion et ses dérivés ainsi que le sulfate (Monaghan *et al.*, 1999). Ce résultat démontre l'importance de l'apport de S aussi bien avant la floraison que pendant le développement du grain.

3.3 L'effet de l'azote et du soufre sur les protéines de réserve du grain

Comme l'N et le S sont tous deux nécessaires à la formation des acides aminés et donc à la synthèse des protéines, le rapport N/S influence la capacité de la plante à synthétiser des protéines plus ou moins riches en S. De nombreuses études ont en effet montré l'impact de la nutrition azotée et soufrée sur la quantité et la composition en protéines du grain de blé. Il est connu qu'un apport d'N à la culture accroît la quantité de PR du grain en augmentant leur durée et leur vitesse d'accumulation (Shewry *et al.*, 2001 ; Triboï *et al.*, 2003 ; Chope *et al.*, 2014). En condition de carence en S, il a été montré une concentration plus faible par grain de cystéine et de méthionine (Wrigley *et al.*, 1980), entraînant une concentration plus faible en PR riches en S, compensée par une concentration plus importante des PR pauvres en S, notamment les HMW-GS (Zhao *et al.*, 1999 ; Wieser *et al.*, 2004). Un apport excédentaire d'N augmente le rapport N/S et entraîne des résultats similaires à une carence en S (Wieser and Seilmeier, 1998 ; Zörb *et al.*, 2010). Il a récemment été démontré avec l'apport de différents traitements N et S à une culture de blé que la quantité des PR pauvres en S dépend directement de la disponibilité

en N du sol, quel que soit le traitement (i.e. la quantité d'N et S apportée), alors que la quantité en PR riches en S dépend de la disponibilité en S du sol (Figure 14 ; Dai *et al.*, 2015).

Le ratio N/S est donc très important pour la mise en place de la qualité du grain et est souvent utilisé pour diagnostiquer le statut soufré dans les cultures (Reussi *et al.*, 2011). Dans le grain de blé, ce ratio varie généralement entre 12 et 25 en fonction des apports de N et S (Randall *et al.*, 1981) et il a été admis qu'un rapport N/S ≥ 17 est synonyme d'une carence en S (Zhao *et al.*, 1999).

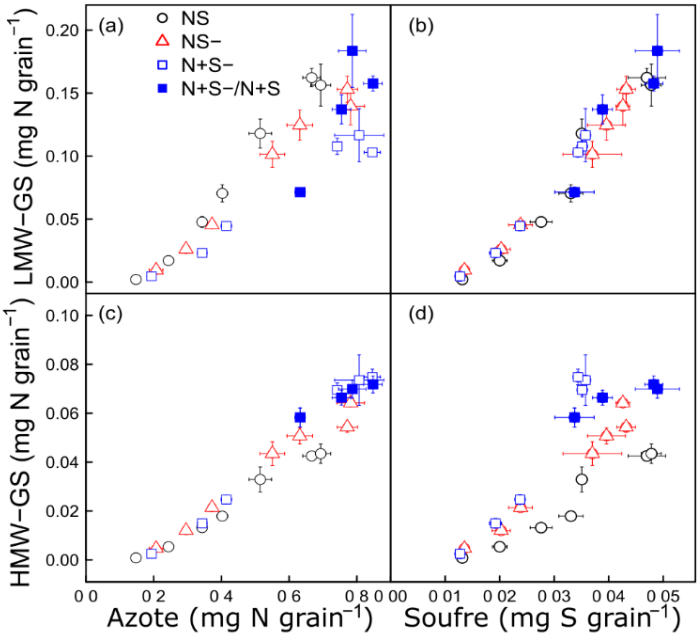


Figure 14. Relations entre la quantité de LMW-GS (a et b) et HMW-GS (c et d) et la quantité d'azote (a et c) et de soufre (b et d) par grain. Les points représentent les moyennes ± erreur standard pour n = 3 répétitions biologiques, pour quatre conditions de nutrition : NS (3 mM N et 2 mM S), NS– (3 mM N et 0.02 mM S), N+S– (15 mM N et 0.02 mM S) et N+S–/N+S (en milieu de remplissage du grain, passage d'une condition de carence en S (N+S–) par une condition de nutrition avec 15 mM N et 2 mM S). D'après Dai *et al.* (2015).

3.4 La qualité du grain dans un contexte de diminution des intrants

Afin d'augmenter les rendements et la production agricole mondiale, l'emploi d'N a considérablement augmenté depuis le milieu du XX^{ème} siècle (Cassman *et al.*, 2003 ; Hirel *et al.*, 2011). La capacité de la plante à capturer l'N du sol dépend du type de sol, de l'environnement et de l'espèce, mais il a été estimé que 50 à 70% de l'N fourni au sol n'est pas fixé par la plante (Hodge *et al.*, 2000). Une partie de cet N est fixée par les microorganismes du sol, mais le surplus peut être lessivé, ce qui entraîne une eutrophisation des eaux présentes dans les écosystèmes à proximité (Giles, 2005 ; Michael Beman *et al.*, 2005). Ceci a pour conséquence un déséquilibre important dans le fonctionnement de ces écosystèmes (Matson *et al.*, 2002 ; Hirel *et al.*, 2007). Face à cette problématique, des mesures ont été prises afin de raisonner les apports d'N à la culture et ainsi limiter leur impact sur l'environnement. Dans ce sens a été instaurée la directive européenne « Directive Nitrates » en 1991. Cette directive s'est traduite par la définition de « zones vulnérables » où sont imposées des pratiques agricoles particulières, avec par exemple l'implantation de bandes enherbées en bord de cours d'eau afin de limiter les risques de ruissellement d'N, ou encore avec la définition de périodes d'interdiction d'épandage. En plus de ce problème environnemental, l'apport d'N à la culture et la fertilisation de manière plus générale représentent un coût élevé. Ainsi, pour la production

de blé tendre, le poste fertilisation représente entre 40 et 50% des charges opérationnelles (Hirel, 2013).

La disponibilité en S du sol est également au centre de préoccupations pour la production agricole. Depuis plusieurs années, des progrès considérables ont été réalisés afin de réduire les émissions industrielles de S, conduisant à une diminution du dépôt du S atmosphérique dans les sols (Eriksen, 2009). En parallèle, l'augmentation des rendements s'accompagne d'un appauvrissement des sols en nutriments, notamment en S (Scherer, 2001). Ainsi, une carence des sols en S est aujourd'hui observée. L'emploi de fertilisants qui contiennent de moins en moins de S accroît un peu plus l'écart entre la disponibilité en N et la disponibilité en S et donc la carence en S, qui impacte à la fois le rendement et la composition du grain (Hawkesford, 2000 ; Hawkesford and DeKok, 2006).

Dans ce contexte de modification des conditions de nutrition de la plante de blé, les paramètres de quantité de protéine et de composition du grain doivent être contrôlés afin de maintenir le taux de protéines et la qualité du grain souhaités.

4. LES PROTÉINES NUCLÉAIRES AU CENTRE DES MÉCANISMES DE RÉGULATION

4.1 La régulation transcriptionnelle de la synthèse des protéines de réserve

La régulation des PR du grain par la disponibilité en N et S semble s'effectuer au niveau du processus de transcription. La régulation transcriptionnelle de la synthèse des PR est en effet connue, aussi bien chez les espèces cultivées du groupe des Angiospermes dicotylédones comme le pois (Gatehouse *et al.*, 1982; Evans *et al.*, 1984) que chez les céréales comme l'orge (Rahman *et al.*, 1983; Giese and Hopp, 1984) ou le blé (Dai *et al.*, 2015), c'est-à-dire que la quantité de protéine dépend directement du niveau de transcrit. Chez le blé, il a été montré que la vitesse d'accumulation des différentes classes de gliadines et sous-unités de gluténines était fortement corrélée au niveau d'expression des gènes correspondants, et ce quelles que soient les quantités d'N et de S appliquées à la culture (Figure 15 ; Dai *et al.*, 2015). Ce résultat démontre l'effet important de cette régulation transcriptionnelle.

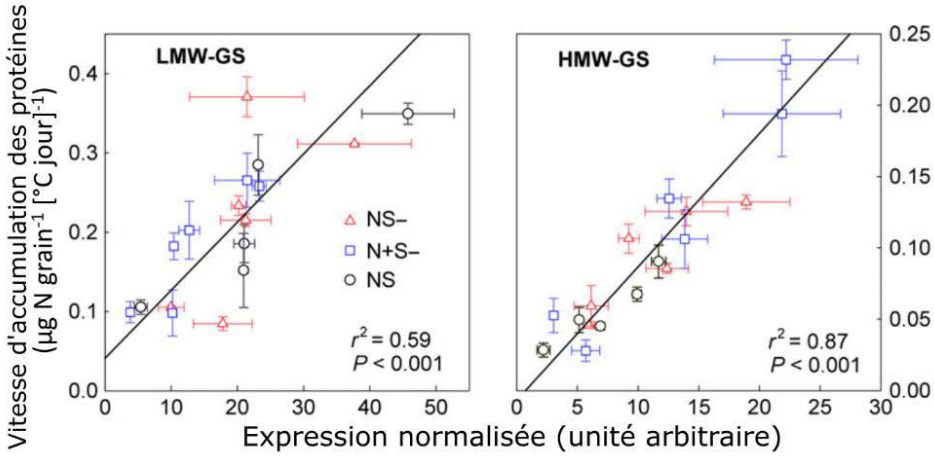


Figure 15. Corrélations entre la vitesse d'accumulation des LMW-GS et HMW-GS et le niveau d'expression des gènes correspondants. Les données correspondent aux moyennes ± erreur standard pour n = 2 répétitions biologiques pour l'expression des gènes et n = 3 répétitions pour les vitesses d'accumulation des PR, pour trois conditions de nutrition : NS (3 mM N et 2 mM S), NS– (3 mM N et 0.02 mM S) et N+S– (15 mM N et 0.02 mM S). D'après Dai *et al.* (2015).

Cette régulation met en jeu plusieurs facteurs de transcription (FT) qui se fixent au niveau de courtes séquences situées dans les régions promotrices des gènes de PR aussi appelées cis-motifs. Un des cis-motifs les plus caractérisés est l'endosperm box, un motif bi-partite constitué de l'élément GCN4-Like Motif (GLM) et de la Prolamin Box (PB). Le premier FT intervenant dans cette régulation a été décrit chez le maïs, il s'agit d'*Opaque2*, qui se fixe sur le motif GLM (Lohmer *et al.*, 1991). Aujourd'hui, huit FT sont connus, tous étant activateurs de la transcription des gènes de PR. Les interactions entre ces FT et les cis-motifs ont fait l'objet de nombreuses études, notamment chez l'orge, conduisant à proposer un schéma de régulation de la synthèse des PR (Diaz *et al.*, 2002; Rubio-Somoza *et al.*, 2006a et b ; Moreno-Risueno *et al.*, 2008). Ce schéma semble être conservé chez les céréales (Verdier and Thompson, 2008 ; Xi and Zheng, 2011) et la fixation de certains FT sur les cis-motifs de gènes de PR a été confirmée chez le blé (Albani *et al.*, 1997 ; Conlan *et al.*, 1999 ; Ravel *et al.*, 2014).

Le noyau apparaît donc comme un lieu de régulation de l'accumulation des PR dans le grain de blé. Il a également été mis en évidence un effet de la nutrition sur ce schéma de régulation. En effet, le motif GLM est activateur en condition de nutrition azotée optimale et devient répresseur en condition de carence en N (Muller and Knudsen, 1993). Malgré ces différents éléments, les mécanismes moléculaires qui régulent l'expression des gènes de PR sont encore peu connus, et les interactions qui surviennent *in vivo* entre les différents acteurs sont probablement à un niveau de complexité plus élevé (Mehrotra *et al.*, 2009).

4.2 Le noyau : un compartiment cellulaire très dynamique

Au sein de la cellule eucaryote, le noyau est le compartiment cellulaire qui renferme l'essentiel de l'information génétique, également contenue au sein des mitochondries et des chloroplastes. L'ADN nucléaire forme avec des protéines appelées histones les unités de base de la chromatine, les nucléosomes. La particule de cœur du nucléosome est formée d'un octamère d'histones (deux exemplaires de chacune des histones H2A, H2B, H3 et H4) autour duquel s'enroule l'ADN. Cette structure est ensuite scellée par les histones H1. C'est le premier niveau de compaction de la chromatine. Cette compaction n'est pas homogène tout au long de la molécule d'ADN. Des zones faiblement condensées, l'euchromatine, accessibles à la machinerie de transcription et des zones fortement condensées et inaccessibles correspondant à l'hétérochromatine sont observées. Des complexes de remodelage de la chromatine sont capables de modifier cet état de compaction et ainsi de réguler la transcription des gènes. Ce mécanisme épigénétique intervient dans la régulation du développement de la plante et dans la réponse de la plante aux facteurs environnementaux (Jarillo *et al.*, 2009).

En plus de contenir la chromatine, le noyau renferme différentes structures (Figure 16 ; Petrovská *et al.*, 2015). La plus connue est le nucléole, dont les principales fonctions sont la synthèse des ARN ribosomiques et l'assemblage des ribosomes (Fromont-Racine *et al.*, 2003; Boisvert *et al.*, 2007). Chez *A. thaliana*, 217 protéines ont été identifiées au sein du nucléole, dont beaucoup correspondent à des protéines ribosomales (Pendle *et al.*, 2005). La traduction ayant lieu dans le cytoplasme, il y a un échange de matériel permanent entre le noyau et le cytoplasme, avec une sortie des ARNm et une importation de protéines qui possèdent un signal de localisation nucléaire. Ce transport nucléo-cytoplasmique est permis par des structures qui interrompent l'enveloppe nucléaire, appelées pores nucléaires (Nuclear Pore Complex, NPC ; Tamura and Hara-Nishimura, 2013) et fait intervenir des protéines de transport : les importines et exportines. L'export des ARNm à travers les NPC est très contrôlé (Erkmann and Kutay, 2004) et il existe une synchronisation entre l'import des protéines ribosomales et le processus d'assemblage des ribosomes (Kressler *et al.*, 2012). Le transport nucléo-cytoplasmique est donc un mécanisme très contrôlé, ayant un rôle central dans le métabolisme cellulaire et le développement de la plante (Meier and Brkljacic, 2009; Merkle, 2011; Tamura and Hara-Nishimura, 2014). Ce transport est également très rapide puisqu'il a été estimé que le nombre de translocations par NPC avoisine les 800 par seconde (Fried and Kutay, 2003). D'autres structures existent au sein du noyau comme les corps de Cajal et les granules (speckles), qui semblent avoir un rôle dans l'épissage des ARN pré-messagers.

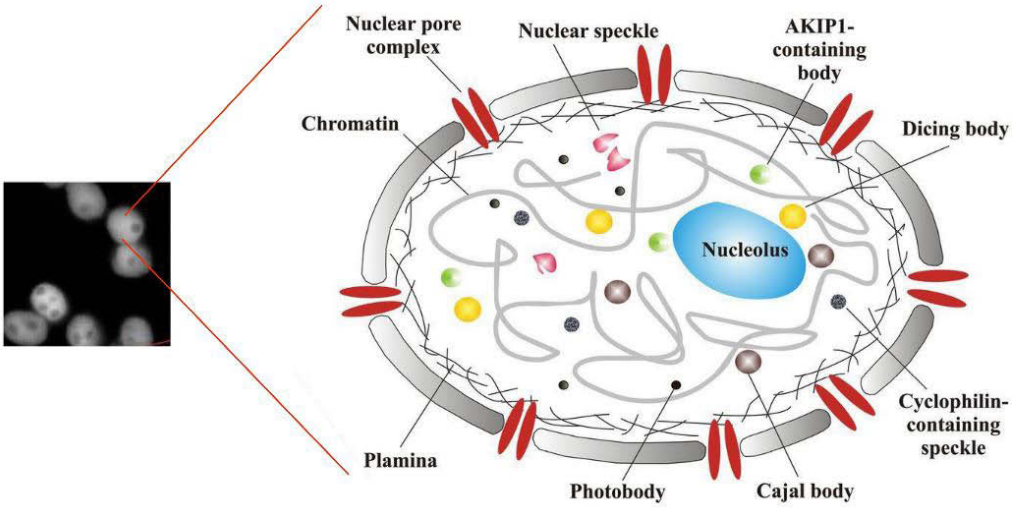


Figure 16. Représentation schématique d'un noyau de cellule végétale. Sur la gauche, une photo de noyaux purifiés par cytométrie de flux, colorés au DAPI et observés en microscopie à fluorescence, chez l'orge. D'après Petrovská *et al.* (2015).

Le noyau est donc une structure très dynamique, renfermant une grande diversité de protéines, parfois présentes dans le noyau de manière très ponctuelle, et intervenant dans des processus très diversifiés. L'ensemble de ces processus, liés de près ou de loin à l'ADN, via sa réplication, sa transcription ou sa conformation, à la traduction ou encore au trafic intracellulaire a un rôle majeur dans le développement de la plante. Bien que la structure et les fonctions du noyau soient assez bien caractérisées, l'état des connaissances est largement dû aux études menées au sein des cellules animales et l'étude du protéome nucléaire chez les plantes en est à ses prémices.

4.3 L'extraction et l'analyse des protéines nucléaires

Chez la souris, les protéines nucléaires représentent environ 14% du protéome total (Fink *et al.*, 2008). Une proportion comparable a été estimée chez les plantes, entre 10 et 20% (Narula *et al.*, 2013), mais la diversité de ces protéines est souvent sous-représentée dans les études de protéomique, notamment du fait de la forte proportion des protéines histones dans les extraits nucléaires. Les protéines nucléaires intervenant dans les mécanismes de régulation sont souvent en très faible abondance, comme par exemple les FT dont une faible quantité est souvent suffisante pour activer ou réprimer l'expression des gènes qu'ils régulent. Pour avoir une vision la plus représentative possible de la diversité protéique présente au sein du noyau, il est ainsi nécessaire d'avoir recours à des étapes de purification de noyaux et/ou d'enrichissement en protéines nucléaires.

Les récents progrès méthodologiques et technologiques en termes de fractionnement subcellulaire et en spectrométrie de masse permettent aujourd'hui d'étudier le protéome au niveau des différents organites de la cellule : vacuole, mitochondrie, chloroplaste, noyau et même au niveau du nucléole comme nous l'avons vu précédemment. Diverses méthodes existent pour isoler et purifier l'organite souhaité, dont les plus courantes sont la centrifugation différentielle et la centrifugation en gradient de densité (Lee *et al.*, 2010). La centrifugation

différentielle consiste en une centrifugation de l'extrait cellulaire à différentes vitesses, chaque vitesse entraînant le dépôt dans le culot d'un organite en particulier. Pour la centrifugation en gradient de densité, l'extrait cellulaire est déposé dans un gradient continu ou discontinu, réalisé avec des milieux de viscosité, d'osmolarité ou de densité différents. Les plus utilisés sont les gradients de sucrose, de Ficoll ou de Percoll. Après centrifugation, l'organite se concentre dans le gradient à l'endroit correspondant à sa propre densité. Cette méthode basée sur un gradient de densité a été la plus utilisée pour purifier des noyaux dans les études menées chez les plantes (Petrovská *et al.*, 2015), et a été utilisée au cours de cette thèse. Récemment chez l'orge, une méthode de purification de noyaux par cytométrie de flux a été développée et a permis l'identification de plus de 800 protéines (Petrovská *et al.*, 2014).

4.4 L'étude des protéines nucléaires chez les plantes

Même si le fonctionnement du noyau de la cellule végétale n'est pas encore totalement élucidé et si les acteurs nucléaires ne sont pas tous identifiés et caractérisés, l'exploration du protéome nucléaire a tout de même fait l'objet de plusieurs études (Figure 17 ; Petrovská *et al.*, 2015). Chez les céréales, des études chez le maïs (Ferreira *et al.*, 2006; Casati *et al.*, 2008; Guo *et al.*, 2014), le riz (Khan and Komatsu, 2004; Tan *et al.*, 2007; Li *et al.*, 2008; Aki and Yanagisawa, 2009; Jaiswal *et al.*, 2013) et récemment l'orge (Petrovská *et al.*, 2014) ont été réalisées. La plupart de ces études ont été effectuées sur des plantules ou à partir de feuilles, et peu finalement se sont intéressées au protéome nucléaire du grain (Repetto *et al.*, 2012). Deux études principales ont analysé l'albumen de grains de riz (Li *et al.*, 2008) et des graines entières de luzerne (Repetto *et al.*, 2008), permettant d'avoir une première vision globale des protéines nucléaires mises en action pendant le développement du grain.

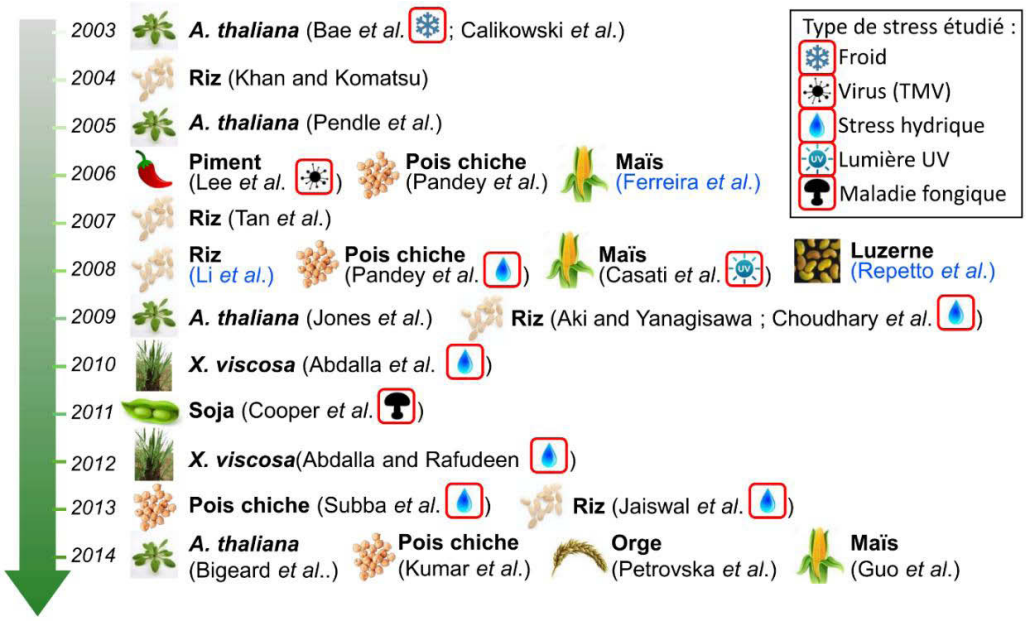


Figure 17. Chronologie des analyses du protéome nucléaire réalisées chez les plantes. Les études indiquées en bleu ont été réalisées à partir de grains. Lorsque le protéome nucléaire a été analysé en réponse à des facteurs environnementaux, la nature du stress appliqué est indiquée par un symbole encadré en rouge. Adapté de Petrovská *et al.* (2015).

Outre l'objectif d'obtenir des informations sur les fonctions du noyau de la cellule végétale, plusieurs des études présentées en Figure 17 ont ciblé le protéome nucléaire afin de mettre en évidence des protéines régulatrices impliquées dans la réponse de la plante à l'environnement. Divers types de facteurs environnementaux ont ainsi été étudiés, notamment le stress hydrique (Pandey *et al.*, 2008 ; Choudhary *et al.*, 2009 ; Abdalla *et al.*, 2010 ; Abdalla and Rafudeen, 2012 ; Subba *et al.*, 2013 ; Jaiswal *et al.*, 2013), le stress au froid (Bae *et al.*, 2003) ou encore le stress dû à une maladie fongique comme la rouille du soja (Cooper *et al.*, 2011). Chez le riz par exemple, l'application d'un stress hydrique a permis de mettre en évidence 78 protéines nucléaires potentiellement au centre de mécanismes d'adaptation de la plante à la sécheresse, comme des FT et des protéines de remodelage de la chromatine (Jaiswal *et al.*, 2013). L'étude du protéome nucléaire est donc très utile pour avoir accès aux protéines régulatrices du développement de la plante et de son adaptation à l'environnement, qui sont souvent diluées dans l'extrait protéique total.

5. L'INTEGRATION DE DONNEES -OMIQUES POUR VISUALISER LA REPONSE BIOLOGIQUE

5.1 Biologie des systèmes

Les évolutions récentes en termes de méthodologie et de technologie permettent aujourd'hui d'obtenir de très nombreuses informations sur la réponse cellulaire à un stress ou un stimulus. L'apparition des technologies de séquençage nouvelle génération a par exemple considérablement augmenté les débits de séquençage d'ADN tout en diminuant les coûts associés et a permis l'émergence de nouvelles approches comme le séquençage massif d'ARN, qui permet de connaître l'intégralité des gènes exprimés dans un échantillon, que le génome de l'espèce étudiée soit complet et disponible ou non (Conesa *et al.*, 2016). Des progrès ont également été réalisés en spectrométrie de masse et permettent aujourd'hui de détecter et quantifier finement une large diversité de protéines et métabolites présents dans un échantillon. Plusieurs échelons sont distingués parmi les données -omiques, avec notamment la transcriptomique, la protéomique et la métabolomique qui font le lien entre le génome et le phénotype observé (Figure 18 ; Mochida and Shinozaki, 2011). De nouveaux types de données -omiques voient également le jour avec par exemple l'étude de l'épigénome qui correspond à l'ensemble des marques épigénétiques du génome ou encore l'hormonome, représentant les molécules hormonales, qui ont un intérêt important chez les plantes du fait de leur rôle dans les mécanismes de signalisation moléculaire et de réponse à l'environnement.

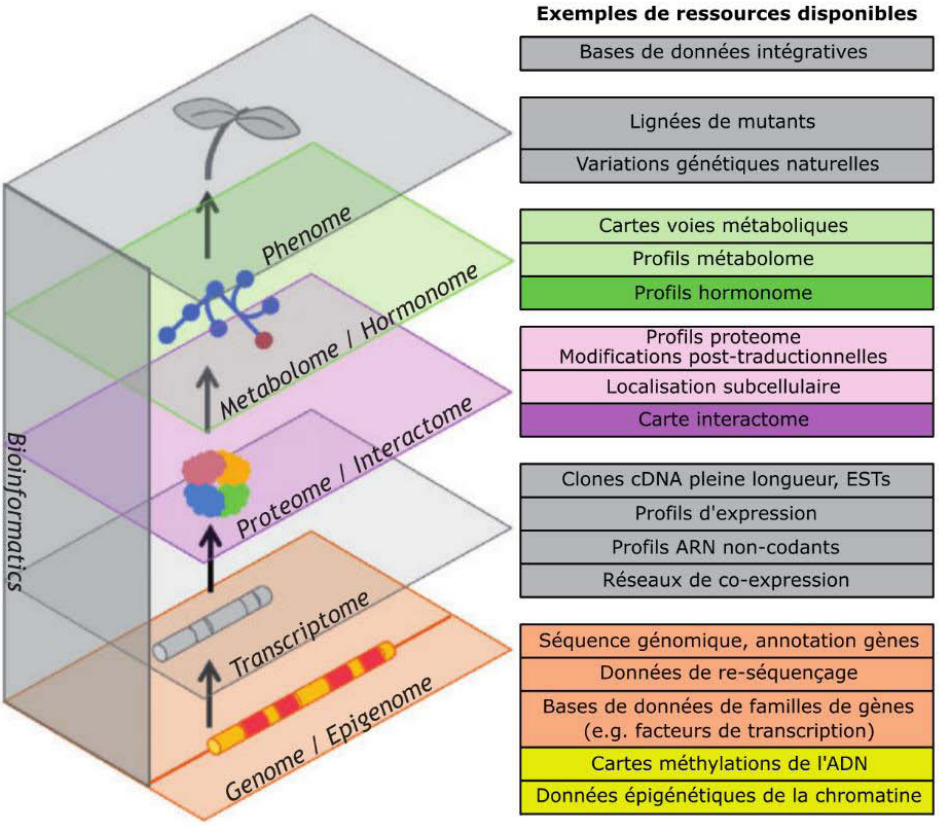


Figure 18. Les différents niveaux de données -omiques et des exemples de ressources disponibles à chaque niveau chez *Arabidopsis thaliana.* D'après Mochida and Shinozaki (2011)

Les technologies -omiques permettent donc d'obtenir de nombreuses informations à différents niveaux biologiques, de la séquence des gènes à l'accumulation des protéines et des composés métaboliques. Ces données peuvent ainsi couvrir tous les mécanismes impliqués dans les variations qui se produisent dans la cellule et qui influencent le fonctionnement du système d'étude. Les technologies -omiques représentent une nouvelle procédure de recherche dans laquelle la formulation et la vérification des hypothèses biologiques sont effectuées en aval de la production des données. Elles sont donc particulièrement utiles pour l'exploration sans *a priori* et la compréhension des mécanismes moléculaires impliqués dans la réponse de la plante, par exemple à un facteur environnemental (Urano *et al.*, 2010).

Avec l'utilisation croissante de ces approches -omiques s'est ainsi développée la biologie des systèmes, concept introduit par Ludwig von Bertalanffy (Von Bertalanffy, 1968). Contrairement à la biologie moléculaire traditionnelle qui se concentre sur la caractérisation de composants individuels de la cellule (ARNs, protéines, métabolites, etc), la biologie des systèmes est de nature intégrative et considère les différents composants comme faisant partie d'un ensemble (Carvunis *et al.*, 2009). Ainsi le principe repose sur la combinaison des données -omiques obtenues à différentes échelles dans le but d'analyser et comprendre comment les interactions observées dans la cellule à ces différents niveaux expliquent le phénotype observé.

Mais face à la multiplication des données générées, la principale difficulté réside dans la capacité de les analyser et les intégrer afin de tirer des conclusions biologiques. En effet, les multiples interactions entre les différents composants du système rendent souvent les interprétations difficiles et témoignent de la réelle complexité de la réponse cellulaire (Fernie, 2012).

5.2 Les différents types de réseaux biologiques

Face à cette accumulation des données -omiques, les réseaux biologiques sont rapidement apparus comme une approche attractive pour extraire de ces larges jeux de données l'information nécessaire à la compréhension des mécanismes moléculaires qui interviennent dans une réponse biologique (Usadel *et al.*, 2009 ; Moreno-Risueno *et al.*, 2010). Graphiquement, les réseaux biologiques sont représentés par des sommets qui peuvent correspondre à diverses entités moléculaires telles que des gènes, protéines ou encore des métabolites, et par des arêtes qui relient les sommets entre eux. Il existe différents types de réseaux en fonction de la signification des arêtes, c'est-à-dire en fonction de la nature de l'interaction qui existe entre les sommets. Ainsi il existe des réseaux de co-expression, d'interactions protéine-protéine, les réseaux métaboliques et les réseaux de régulation et de signalisation (Figure 19 ; Bassel *et al.*, 2012).

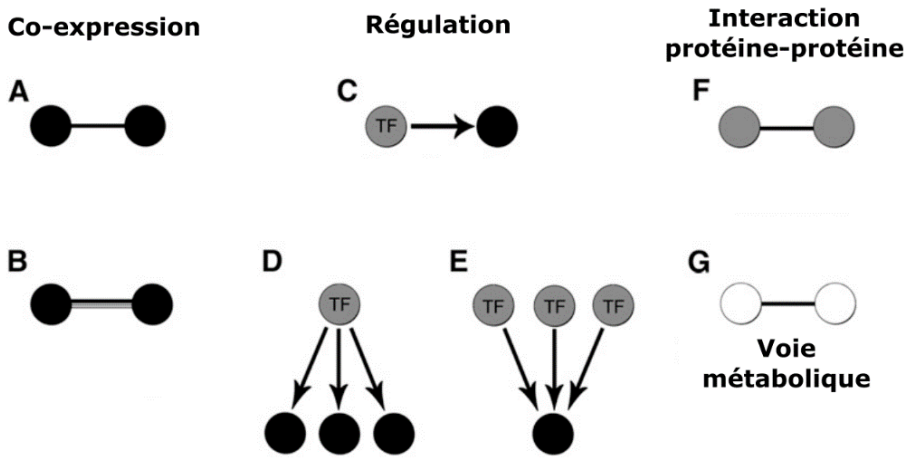


Figure 19. Représentation schématique de différents types de réseaux biologiques. (A) et (B) Réseaux de co-expression, deux gènes sont liés s'ils ont un profil d'expression similaire (A) et/ou partagent des propriétés communes (B). (C), (D) et (E) Réseaux de régulation, liaison entre un FT et le promoteur d'un (C) ou plusieurs (D) gène(s) qu'il régule, liaison de plusieurs FT sur le même promoteur d'un gène (E). (F) Réseau d'interaction physique protéine-protéine démontrée expérimentalement. (G) Réseau d'interaction métabolique représentant une réaction biochimique qui convertit un métabolite en un autre. Les gènes sont colorés en noir, les protéines en gris et les métabolites en blanc. D'après Bassel *et al.* (2012).

Les réseaux de co-expression se sont fortement développés ces dernières années avec l'obtention de grands jeux de données dans le domaine de la transcriptomique, générés avec les approches de puce à ADN et de RNA-Seq (Serin *et al.*, 2016). Pour ce type de réseau, les profils d'expression des différents gènes sont comparés et un score de similarité est calculé pour chaque comparaison de deux gènes, par exemple *via* le calcul d'un coefficient de corrélation. A partir d'un certain seuil, on considère que les gènes sont co-exprimés. Le réseau est alors construit à partir de la liste des gènes co-exprimés. Dans ce type de réseau, le principe de « guilt by association » est appliqué, c'est-à-dire que les gènes qui ont des profils d'expression similaires sont susceptibles de partager des fonctions similaires ou de faire partie de la même voie de régulation (Oliver, 2000). Ainsi, au sein des groupes de gènes fortement interconnectés dans le réseau, appelés modules, il est possible à partir des fonctions qui y sont représentées d'émettre des hypothèses sur les fonctions des gènes non caractérisés fonctionnellement (Aoki *et al.*, 2007; Silva *et al.*, 2016). Les modules sont souvent utilisés comme point de départ dans l'interprétation puisqu'ils réduisent considérablement la complexité du réseau (Aoki *et al.*, 2007). Les sommets qui apparaissent avec un fort degré de connexion sont alors identifiés comme des « hubs » au sein des modules et sont souvent considérés comme des gènes essentiels du réseau (Provero, 2002).

Les réseaux d'interactions protéine-protéine ou d'interactions protéine-ADN sont construits à partir de résultats issus des méthodes de double hybride (interaction protéine-protéine) par exemple et de simple hybride ou d'immunoprécipitation de chromatine (interaction protéine-ADN). Dans les réseaux générés, les interactions correspondent ainsi à des liens physiques qui ont été démontrés expérimentalement. Dans cette optique, de nombreux réseaux sont générés pour étudier les interactions entre des FT et des séquences ADN (Gaudinier and Brady, 2016). Chez *A. thaliana* par exemple, des expériences de simple hybride ont été à l'origine d'un réseau transcriptionnel qui a permis de révéler de nouveaux régulateurs

de la synthèse de la paroi secondaire des cellules racinaires, modulés par le stress salin ou la carence en fer (Taylor-Teeples *et al.*, 2014). Par rapport à un réseau de co-expression qui est non orienté, le réseau de régulation transcriptionnel généré dans cette étude va à un niveau plus loin dans l'analyse puisqu'il permet de caractériser le sens et le type d'interaction entre les deux entités moléculaires (Figure 19).

Il est également possible d'apporter les informations d'interactions physiques ou de régulation disponibles pour les gènes, protéines ou métabolites présents dans un réseau de co-expression afin d'avoir une vision la plus complète possible. Pour faire le lien entre le génotype et le phénotype et comprendre les mécanismes de régulation impliqués il apparaît en effet nécessaire dans la mesure du possible de combiner les réseaux obtenus à différents niveaux moléculaires (gènes, protéines, métabolites ; Gaudinier *et al.*, 2015).

5.3 RulNet : l'outil utilisé au cours de la thèse

Plusieurs méthodes existent pour inférer des réseaux biologiques, en fonction des données disponibles ou encore de la nature des interactions. Dans cette thèse, des données de quantités de protéines et de métabolites et des données d'expression de gènes ont été obtenues. La construction de réseaux de co-expression a donc pu être envisagée pour intégrer les données générées. Plusieurs méthodes sont couramment utilisées pour générer ce type de réseaux. En fonction de la méthode, il existe ainsi les réseaux booléens, les Relevance Networks, les réseaux bayésiens et les réseaux basés sur la découverte de règles d'association.

Pour les réseaux booléens, les données d'expression sont binarisées, un gène est soit dans un état actif soit inactif. L'objectif est de trouver des règles logiques permettant de déterminer l'état d'un gène à un instant t+1 lorsque l'on connait l'état de ce gène et d'autres gènes à un instant t. Les réseaux bayésiens se basent quant à eux sur des distributions de probabilités conditionnelles entre les variables. Pour chaque couple de gènes (G1, G2), la probabilité d'avoir G1 sachant G2 est calculée. Une relation de causalité existe alors si cette probabilité est importante. Dans le cas des Relevance Networks, la corrélation entre chaque paire de gènes est calculée, basée sur le coefficient de corrélation de Pearson. Les coefficients sont ensuite filtrés et conservés s'ils dépassent un certain seuil.

L'outil RulNet (http://rulnet.isima.fr/ ; Vincent *et al.*, 2015), utilisé au cours de cette thèse pour générer des réseaux de co-expression, est basé sur la découverte de règles d'association. Une règle est une expression de la forme X → Y et se lit « X implique Y ». Ainsi cela signifie que lorsqu'une propriété, appelée prédicat est observée sur X (antécédent) alors un prédicat est observé sur Y (conséquent), les prédicats étant identiques ou non. Il existe ensuite différentes mesures de qualité des règles qui permettent d'évaluer leur pertinence, dont le support, la confiance et le lift. Le support exprime la fréquence d'une règle, et correspond à la probabilité que X et Y soient simultanément satisfaits (exemple : proportion d'échantillons dans lesquels le gène 1 et le gène 2 sont tous les deux surexprimés). La confiance correspond à la probabilité que Y soit satisfait sachant que X est satisfait. Lorsque la confiance est égale à 1 la règle est dite exacte, sinon elle est approximative. Le lift mesure quant à lui la dépendance entre X et Y

et correspond au rapport entre la probabilité réelle d'avoir X et Y satisfaits et la probabilité attendue si X et Y étaient statistiquement indépendants.

La méthode de découverte de règles d'association a été pour la première fois appliquée à l'inférence de réseaux de gènes en 1997 (Brazma *et al.*, 1997). Avec RulNet, cette approche a été adaptée à l'intégration de données -omiques chez le blé (Vincent *et al.*, 2015 ; Dai *et al.*, 2015). Ainsi, X et Y peuvent correspondre à différentes variables comme des gènes, des ARNs, protéines, métabolites ou encore des variables phénotypiques. Dans RulNet, les sémantiques choisies en fonction des règles à inférer sont rédigées sous la forme de requêtes RQL (Chardin *et al.*, 2014). Il est possible de définir des seuils pour les différentes mesures de qualité des règles mentionnées précédemment, afin de ne générer que les règles qui seront considérées comme significatives. Le réseau est ensuite visualisable sous d'autres outils comme Cytoscape (http://www.cytoscape.org/ ; Smoot *et al.*, 2011).

Cette approche laisse la possibilité au biologiste de générer plusieurs réseaux ou un seul réseau global à partir de différentes sémantiques qu'il peut choisir en fonction de ses objectifs biologiques et des caractéristiques de ses données. L'utilisateur a ainsi le choix de la signification biologique des liens à inférer. Il est également possible de définir des gènes centraux, c'est-à-dire des gènes qu'il souhaite voir apparaître dans le réseau, et les règles générées seront alors limitées à celles impliquant ces entités d'intérêt.

6. CONCLUSIONS DE LA SYNTHESE ET STRATEGIE D'ETUDE

Comme nous avons pu le voir dans cette synthèse bibliographique, la qualité du grain de blé est essentiellement déterminée par la teneur et la composition en PR. L'accumulation des différentes classes de gliadines et sous-unités de gluténines, qui dépend fortement de la disponibilité en N est S du sol, doit être contrôlée afin d'obtenir les caractéristiques nutritionnelles et technologiques souhaitées. Or, le contexte actuel avec un objectif d'augmentation du rendement en grains et des conditions de nutrition de la plante modifiées, avec une limitation des apports d'engrais azotés et une carence en S observée dans les sols, menace de réduire la qualité du grain. Afin de maintenir cette qualité, il est nécessaire de comprendre les mécanismes moléculaires impliqués dans la régulation du développement du grain et dans la régulation de la synthèse des PR en réponse à la nutrition azotée et soufrée. Il apparaît également que les PR du grain de blé sont régulées au niveau transcriptionnel. Le noyau apparait donc comme un compartiment cellulaire très important et une cible de choix pour l'étude de ces mécanismes de régulation. Le but de cette thèse a ainsi été d'étudier le protéome nucléaire du grain de blé au cours du développement et en réponse à différents apports de N et S, puis d'obtenir une vision globale de la réponse du grain à la nutrition dans une approche intégrative combinant différents types de données -omiques (transcrits, métabolites, protéines). L'étude a été divisée en trois objectifs principaux (Figure 20).

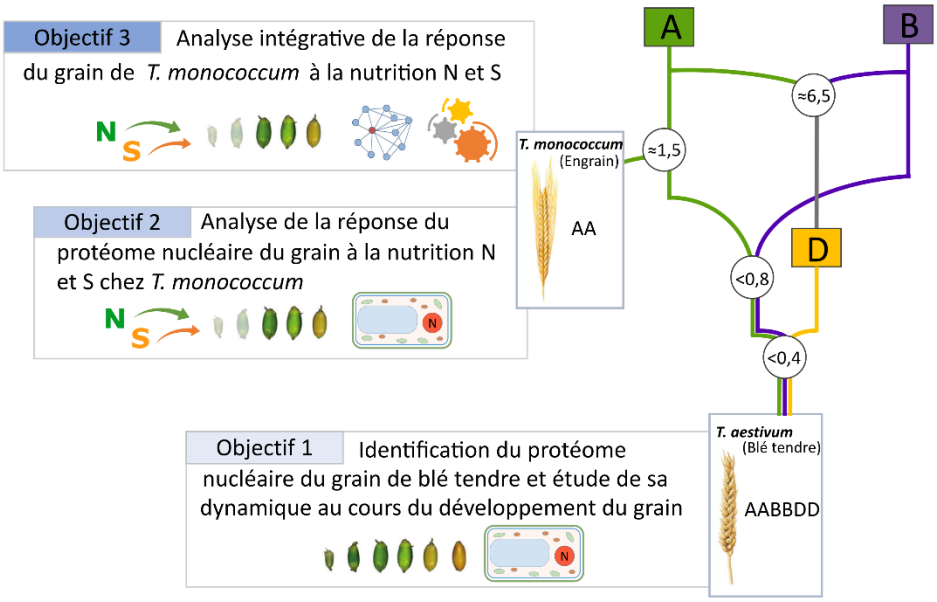


Figure 20. Schéma situant les objectifs du projet de thèse en lien avec les espèces de blé étudiées.

(1) Basé sur une méthode de purification de noyaux et d'extraction de protéines nucléaires mise au point par Emmanuelle Bancel avant mon arrivée qui a dû être validée (**Chapitre 2**), le premier objectif a consisté à identifier le protéome nucléaire du grain de blé tendre (*T. aestivum*) jusqu'alors inexploré, puis étudier les variations d'abondance de ces protéines au cours du développement du grain (**Chapitre 3**).

(2) Le second objectif a été d'identifier les protéines nucléaires qui pouvaient être impactées par la nutrition azotée et soufrée pendant le remplissage du grain et de lier ces variations avec l'accumulation des PR et des protéines métaboliques (**Chapitre 4**). Ces analyses

protéomiques ont été réalisées chez l'engrain (*T. monococcum*), espèce diploïde choisie comme modèle du blé tendre.

(3) Le troisième objectif consistait à obtenir une vision intégrative de la réponse de *T. monococcum* à la nutrition azotée et soufrée (**Chapitre 5**). Pour cela, des analyses de transcriptomique et de métabolites ont été réalisées. L'ensemble des données obtenues à partir du même matériel biologique (données de transcripomique, métabolomique et enfin de protéomique analysées dans l'objectif 2) a été intégré au sein de réseaux biologiques.

REFERENCES

Abdalla, K.O., Baker, B. and Rafudeen, M.S. (2010) Proteomic analysis of nuclear proteins during dehydration of the resurrection plant *Xerophyta viscosa*. *Plant Growth Regul.*, **62**, 279–292.

Abdalla, K.O. and Rafudeen, M.S. (2012) Analysis of the nuclear proteome of the resurrection plant *Xerophyta viscosa* in response to dehydration stress using iTRAQ with 2DLC and tandem mass spectrometry. *J. Proteomics*, **75**, 2361–74.

Acevedo, E., Silva, P. and Silva, H. (2002) Wheat growth and physiology. Available at: http://www.fao.org/3/a-y4011e/y4011e06.htm.

Aki, T. and Yanagisawa, S. (2009) Application of rice nuclear proteome analysis to the identification of evolutionarily conserved and glucose-responsive nuclear proteins. *J. Proteome Res.*, **8**, 3912–24.

Albani, D., Hammond-Kosack, M.C.U., Smith, C., Conlan, S., Colot, V., Holdsworth, M. and Bevan, M.W. (1997) The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. *Plant Cell*, **9**, 171–184.

Altenbach, S.B., Kothari, K.M. and Lieu, D. (2002) Environmental conditions during wheat grain development alter temporal regulation of major gluten protein genes. *Cereal Chem.*, **79**, 279–285.

Aoki, K., Ogata, Y. and Shibata, D. (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.*, **48**, 381–390.

Arvalis-Institut du végétal (2016) Comment s'élabore le rendement des céréales à paille ? Available at: http://www.arvalis-infos.fr/comment-s-elabore-le-rendement--@/view-16191-arvarticle.html.

Asseng, S., Ewert, F., Martre, P., et al. (2015) Rising temperatures reduce global wheat production. *Nat. Clim. Chang.*, **5**, 143–147.

Austin, R.B., Ford, M.A., Edrich, J.A. and Blackwell, R.D. (1977) The nitrogen economy of winter wheat. *J. Agric. Sci.*, **88**, 159–167.

Bae, M.S., Cho, E.J., Choi, E.-Y. and Park, O.K. (2003) Analysis of the *Arabidopsis* nuclear proteome and its response to cold stress. *Plant J.*, **36**, 652–663.

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

Bassel, G.W., Gaudinier, A., Brady, S.M., Hennig, L., Rhee, S.Y. and Smet, I. De (2012) Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. *Plant Cell*, **24**, 3859–3875.

Bates, B., Lennox, A., Prentice, A., Bates, C., Page, P., Nicholson, S. and Swan, G. (2011) National diet and nutrition survey : results from years 1-4 (combined) of the rolling programme about public health england. *Public Heal. England, London, UK*, **4**, 1–24.

Bertalanffy, L. Von (1968) *General system theory*, Georg. Braziller New York.

Bogard, M., Allard, V., Brancourt-Hulmel, M., Heumez, E., Machet, J.-M., Jeuffroy, M.-H., Gate, P., Martre, P. and LeGouis, J. (2010) Deviation from the grain protein concentration-grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *J. Exp. Bot.*, **61**, 4303–4312.

Boisvert, F.-M., Koningsbruggen, S. van, Navascués, J. and Lamond, A.I. (2007) The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.*, **8**, 574–85.

Bonjean, A. (1994) Histoire de la culture des céréales et en particulier de celle du blé tendre (*Triticum aestivum* L.). *Doss. l'environnement l'INRA*, **21**, 29–37.

Brancourt-Hulmel, M., Doussinault, G., Lecomte, C., Bérard, P., Buanec, B. Le and Trottet, M. (2003) Genetic improvement of agronomic traits of winter wheat cultivars

released in France from 1946 to 1992. *Crop Sci.*, **43**, 37–45.

Branlard, G., Dardevet, M., Saccomano, R., Lagoutte, F. and Gourdon, J. (2001) Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica*, **119**, 59–67.

Branlard, G., Lesage, V.S., Bancel, E., Martre, P., Méleard, B. and Rhazi, L. (2015) Coping with wheat quality in a changing environment: proteomics evidence for stress caused by environmental changes. In *Advances in Wheat Genetics: From Genome to Field*. pp. 255–264.

Brazma, A., Vilo, J., Ukkonen, E. and Valtonen, K. (1997) Data mining for regulatory elements in yeast genome. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **5**, 65–74.

Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X. and Huard, F. (2010) Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *F. Crop. Res.*, **119**, 201–212.

Buchner, P. and Hawkesford, M.J. (2014) Complex phylogeny and gene expression patterns of members of the nitrate transporter 1/peptide transporter family (NPF) in wheat. *J. Exp. Bot.*, **65**, 5697–5710.

Buchner, P., Parmar, S., Kriegel, A., Carpentier, M. and Hawkesford, M.J. (2010) The sulfate transporter family in wheat: tissue-specific gene expression in relation to nutrition. *Mol. Plant*, **3**, 374–389.

Carvunis, A., Gomez, E., Thierry-mieg, N., Trilling, L. and Vidal, M. (2009) Biologie systémique. *Medecine/Sciences*, **25**, 578–584.

Casati, P., Campi, M., Chu, F., Suzuki, N., Maltby, D., Guan, S., Burlingame, A.L. and Walbot, V. (2008) Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. *Plant Cell*, **20**, 827–42.

Cassman, K.G., Dobermann, A., Walters, D.T. and Yang, H. (2003) Meeting cereal demand while protecting natural resources and improving environmental quality. *Annu. Rev. Environ. Resour.*, **28**, 315–358.

Chardin, B., Coquery, E., Pailloux, M. and Petit, J.-M. (2014) RQL: a SQL-like query language for discovering meaningful rules. *Int. Conf. data Min. Work.*, **1**, 1203–1206.

Chojecki, A.J.S., Bayliss, M.W. and Gale, M.D. (1986) Cell production and DNA accumulation in the wheat endosperm , and their association with grain weight. *Ann. Bot.*, **58**, 809–817.

Chope, G.A., Wan, Y., Penson, S.P., Bhandari, D.G., Powers, S.J., Shewry, P.R. and Hawkesford, M.J. (2014) Effects of genotype, season, and nitrogen nutrition on gene expression and protein accumulation in wheat grain. *J. Agric. Food Chem.*, **62**, 4399–4407.

Choudhary, M.K., Basu, D., Datta, A., Chakraborty, N. and Chakraborty, S. (2009) Dehydration-responsive nuclear proteome of rice (*Oryza sativa* L.) illustrates protein network, novel regulators of cellular adaptation, and evolutionary perspective. *Mol. Cell. Proteomics*, **8**, 1579–1598.

Conesa, A., Madrigal, P., Tarazona, S., et al. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol.*, **17**, 1–19.

Conlan, R.S., Hammond-Kosack, M. and Bevan, M. (1999) Transcription activation mediated by the bZIP factor SPA on the endosperm box is modulated by ESBF-1 in vitro. *Plant J.*, **19**, 173–181.

Cooper, B., Campbell, K.B., Feng, J., Garrett, W.M. and Frederick, R. (2011) Nuclear proteomic changes linked to soybean rust resistance. *Mol. Biosyst.*, **7**, 773–83.

Corbellini, M., Empilli, S., Vaccino, P., Brandolini, A., Borghi, B., Heun, M. and Salamini, F. (1999) Einkorn characterization for bread and cookie production in relation to protein subunit composition. *Cereal Chem.*, **76**, 727–733.

D'Ovidio, R. and Masci, S. (2004) The low-molecular-weight glutenin subunits of wheat gluten. *J. Cereal Sci.*, **39**, 321–339.

Dai, Z., Plessis, A., Vincent, J., et al. (2015) Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. *Plant J.*, **83**, 326–343.

Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martinez, M., Isabel-La Moneda, I. and Carbonero, P. (2002) The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *Plant J.*, **29**, 453–464.

Dubcovsky, J. and Dvorak, J. (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, **316**, 1862–1866.

Dupont, F.M. and Altenbach, S.B. (2003) Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *J. Cereal Sci.*, **38**, 133–146.

Eriksen, J. (2009) Soil sulfur cycling in temperate agricultural systems. In *Advances in Agronomy*. pp. 55–89.

Erkmann, J. a and Kutay, U. (2004) Nuclear export of mRNA: from the site of transcription to the cytoplasm. *Exp. Cell Res.*, **296**, 12–20.

Evans, I.M., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) Regulation of the transcription of storage-protein mRNA in nuclei isolated from developing pea (*Pisum sativum* L.) cotyledons. *Planta*, **160**, 559–568.

Evers, A.D., Blakeney, A.B. and O'Brien, L. (1999) Cereal structure and composition. *Aust. J. Agric. Res.*, **50**, 629–650.

Evers, T. and Millar, S. (2002) Cereal grain structure and development: some implications for quality. *J. Cereal Sci.*, **36**, 261–284.

FAO (2016) Crop Prospects and Food Situation. *October*, 1–42.

Fazili, I.S., Jamal, A., Ahmad, S., Masoodi, M., Khan, J.S. and Abdin, M.Z. (2008) Interactive effect of sulfur and nitrogen on nitrogen accumulation and harvest in oilseed crops differing in nitrogen assimilation potential. *J. Plant Nutr.*, **31**, 1203–1220.

Feil, B. (1997) The inverse yield-protein relationship in cereals: Possibilities and limitations for genetically improving the grain protein yield. *Trends Agron.*, **1**, 103–119.

Feillet, P. (2000) *Le grain de blé : composition et utilisation* INRA., Paris.

Fernie, A.R. (2012) Grand challenges in plant systems biology: closing the circle(s). *Front. Plant Sci.*, **3**, 35.

Ferreira, B.S., Branco, A.T., Oliveira, M.A. de, Pereira, M.G. and Souza Filho, G.A. de (2006) Methodological improvements on extraction of nuclear proteins and its preliminary analysis during the maize (*Zea mays* L.) endosperm development. *Protein Pept. Lett.*, **13**, 981–4.

Fink, J.L., Karunaratne, S., Mittal, A., et al. (2008) Towards defining the nuclear proteome. *Genome Biol.*, **9**, R15.

FranceAgrimer (2015) Bilans céréaliers : campagne 2014/2015.

Fried, H. and Kutay, U. (2003) Nucleocytoplasmic transport: taking an inventory. *Cell. Mol. Life Sci.*, **60**, 1659–1688.

Fromont-Racine, M., Senger, B., Saveanu, C. and Fasiolo, F. (2003) Ribosome assembly in eukaryotes. *Gene*, **313**, 17–42.

Gatehouse, J.A., Evans, I.M., Bown, D., Croy, R.R.D. and Boulter, D. (1982) Control of storage-protein synthesis during seed development in pea (*Pisum sativum* L.). *Biochem. J.*, **208**, 119–127.

Gaudinier, A. and Brady, S.M. (2016) Mapping transcriptional networks in plants: data-driven discovery of novel biological mechanisms. *Annu. Rev. Plant Biol.*, **67**, 575–94.

Gaudinier, A., Tang, M. and Kliebenstein, D.J. (2015) Transcriptional networks governing plant metabolism. *Curr. Plant Biol.*, **3-4**, 56–64.

Giese, H. and Hopp, H.E. (1984) Influence of nitrogen nutrition on the amount of hordein, protein Z and β-amylase messenger RNA in developing endosperms of barley. *Carlsberg Res. Commun.*, **49**, 365–383.

Giles, J. (2005) Nitrogen study fertilizes fears of pollution. *Nature*, **433**, 791–791.

Giordano, M. and Raven, J.A. (2014) Nitrogen and sulfur assimilation in plants and algae. *Aquat. Bot.*, **118**, 45–61.

Girin, T., Lejay, L., Wirth, J., Widiez, T., Palenchar, P.M., Nazoa, P., Touraine, B., Gojon, A. and Lepetit, M. (2007) Identification of a 150 bp cis -acting element of the AtNRT2.1 promoter involved in the regulation of gene expression by the N and C status of the plant. *Plant. Cell Environ.*, **30**, 1366–1380.

Graybosch, R.A. and Peterson, C.J. (2010) Genetic improvement in winter wheat yields in the great plains of north America, 1959–2008. *Crop Sci.*, **50**, 1882.

Guo, B., Chen, Y., Li, C., et al. (2014) Maize (*Zea mays* L.) seedling leaf nuclear proteome and differentially expressed proteins between a hybrid and its parental lines. *Proteomics*, **14**, 1071–87.

Hawkesford, M.J. (2000) Plant responses to sulphur deficiency and the genetic manipulation of sulphate transporters to improve S-utilization efficiency. *J. Exp. Bot.*, **51**, 131–138.

Hawkesford, M.J. (2003) Transporter gene families in plants: the sulphate transporter gene family - redundancy or specialization? *Physiol. Plant.*, **117**, 155–163.

Hawkesford, M.J. and DeKok, L.J. (2006) Managing sulphur metabolism in plants. *Plant, Cell Environ.*, **29**, 382–395.

Heun, M. (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. *Science*, **278**, 1312–1314.

Hidalgo, A. and Brandolini, A. (2014) Nutritional properties of einkorn wheat (*Triticum monococcum* L.). *J. Sci. Food Agric.*, **94**, 601–612.

Hidalgo, A., Brandolini, A., Pompei, C. and Piscozzi, R. (2006) Carotenoids and tocols of einkorn wheat (*Triticum monococcum* ssp. *monococcum* L.). *J. Cereal Sci.*, **44**, 182–193.

Hirase, K. and Molin, W.T. (2003) Sulfur assimilation in plants and weed control: Potential targets for novel herbicides and action sites of certain safeners. *Weed Biol. Manag.*, **3**, 147–157.

Hirel, B. (2013) *Améliorer l'efficacité d'utilisation de l'azote chez les plantes cultivées*, INRA-Versailles.

Hirel, B., LeGouis, J., Ney, B. and Gallais, A. (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.*, **58**, 2369–2387.

Hirel, B., Tétu, T., Lea, P.J. and Dubois, F. (2011) Improving nitrogen use efficiency in crops for sustainable agriculture. *Sustainability*, **3**, 1452–1485.

Hodge, A., Robinson, D. and Fitter, A. (2000) Are microorganisms more effective than plants at competing for nitrogen? *Trends Plant Sci.*, **5**, 304–308.

Jaiswal, D.K., Ray, D., Choudhary, M.K., et al. (2013) Comparative proteomics of dehydration response in the rice nucleus: new insights into the molecular basis of genotype-specific adaptation. *Proteomics*, **13**, 3478–97.

Jamal, A., Moon, Y.S. and Abdin, M.Z. (2010) Sulphur -a general overview and interaction with nitrogen. *Aust. J. Crop Sci.*, **4**, 523–529.

Jarillo, J.A., Pineiro, M., Cubas, P. and Martinez-Zapater, J.M. (2009) Chromatin remodeling in plant development. *Int. J. Dev. Biol.*, **53**, 1581–1596.

Kataoka, T. (2004) Vacuolar sulfate transporters are essential determinants controlling internal distribution of sulfate in *Arabidopsis*. *Pant Cell*, **16**, 2693–2704.

Khan, M.M.K. and Komatsu, S. (2004) Rice proteomics: recent developments and analysis of nuclear proteins. *Phytochemistry*, **65**, 1671–81.

Kichey, T., Hirel, B., Heumez, E., Dubois, F. and LeGouis, J. (2007) In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. *F. Crop. Res.*, **102**, 22–32.

Kopriva, S. and Rennenberg, H. (2004) Control of sulphate assimilation and glutathione synthesis: interaction with N and C metabolism. *J. Exp. Bot.*, **55**, 1831–1842.

Koprivova, A., Suter, M., Camp, R.O. Den, Brunold, C. and Kopriva, S. (2000) Regulation of sulfate assimilation by nitrogen in Arabidopsis. *Plant Physiol.*, **122**, 737–746.

Kressler, D., Bange, G., Ogawa, Y., et al. (2012) Synchronizing nuclear import of ribosomal proteins with ribosome assembly. *Science*, **338**, 666–71.

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

Lee, Y.H., Tan, H.T. and Chung, M.C.M. (2010) Subcellular fractionation methods and strategies for proteomics. *Proteomics*, **10**, 3935–3956.

Lev-Yadun, S., Gopher, A. and Abbo, A. (2000) The cradle of agriculture. *Science*, **288**, 1602–1603.

Li, G., Nallamilli, B.R.R., Tan, F. and Peng, Z. (2008) Removal of high-abundance proteins for nuclear subproteome studies in rice (*Oryza sativa*) endosperm. *Electrophoresis*, **29**, 604–17.

Lohmer, S., Maddaloni, M., Motto, M., Hartings, H., Salamini, F. and Thompson, R.D. (1991) The maize regulatory locus Opaque-2 encodes a DNA-binding protein which activates the transcription of the b-32 gene. *EMBO*, **10**, 617–624.

Majoul-Haddad, T., Bancel, E., Martre, P., Triboi, E. and Branlard, G. (2013) Effect of short heat shocks applied during grain development on wheat (*Triticum aestivum* L.) grain proteome. *J. Cereal Sci.*, **57**, 486–495.

Malagoli, P., Laine, P., Rossato, L. and Ourry, A. (2005) Dynamics of nitrogen uptake and mobilization in field-grown winter oilseed rape (*Brassica napus*) from stem extension to harvest. II. An 15N-labelling-based simulation model of N partitioning between vegetative and reproductive tissues. *Ann. Bot.*, **95**, 1187–1198.

Malik, a. H. (2009) Nutrient uptake, transport and translocation in cereals: influences of environmental and farming conditions. *Introd. Pap. Fac. Landsc. Planning, Hortic. Agric. Sci.*, **1**, 1–46.

Marcussen, T., Sandve, S.R., Heier, L., et al. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, **345**, 1250092–1250092.

Mares, D., Norstog, K. and Stone, B. (1975) Early stages in the development of wheat endosperm. I. The change from free nuclear to cellular endosperm. *Aust. J. Bot.*, **23**, 311.

Martre, P., Porter, J.R., Jamieson, P.D. and Triboï, E. (2003) Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. *Plant Physiol.*, **133**, 1959–1967.

Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L. and Suzuki, A. (2010) Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. *Ann. Bot.*, **105**, 1141–1157.

Matson, P., Lohse, K.A. and Hall, S.J. (2002) The globalization of nitrogen deposition: consequences for terrestrial ecosystems. *AMBIO A J. Hum. Environ.*, **31**, 113.

Mehrotra, R., Kumar, S., Mehrotra, S. and Singh, B.D. (2009) Seed storage protein gene regulation- A jig-saw puzzle. *Indian J. Biotechnol.*, **8**, 147–158.

Meier, I. and Brkljacic, J. (2009) The nuclear pore and plant development. *Curr. Opin. Plant Biol.*, **12**, 87–95.

Merkle, T. (2011) Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant Cell*

Rep., **30**, 153–176.

Metakovsky, E. V., Chernakov, V.M., Upelniek, V.P., Redaelli, R., Dardevet, M., Branlard, G. and Pogna, N.E. (1996) Recombination mapping of minor omega-gliadin-coding loci on chromosome 1A of common wheat: A revision. *J. Genet. Breed.*, **50**, 277–286.

Michael Beman, J., Arrigo, K.R. and Matson, P.A. (2005) Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature*, **434**, 211–214.

Miflin, B.J. and Habash, D.Z. (2002) The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. *J. Exp. Bot.*, **53**, 979–987.

Miura, K. (2013) Nitrogen and phosphorus nutrition under salinity stress. In *Ecophysiology and Responses of Plants under Salt Stress*. New York, NY: Springer New York, pp. 425–441.

Mochida, K. and Shinozaki, K. (2011) Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol.*, **52**, 2017–2038.

Monaghan, J.M., Scrimgeour, C.M., Stein, W.M., Zhao, F.J. and Evans, E.J. (1999) Sulphur accumulation and redistribution in wheat (*Triticum aestivum*): a study using stable sulphur isotope ratios as a tracer system. *Plant, Cell Environ.*, **22**, 831–839.

Monaghan, J.M., Snape, J.W., Chojecki, A.J.S. and Kettlewell, P.S. (2001) The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield. *Euphytica*, **122**, 309–317.

Monfreda, C., Ramankutty, N. and Foley, J.A. (2008) Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem. Cycles*, **22**, 1–19.

Moreno-Risueno, M.A., Busch, W. and Benfey, P.N. (2010) Omics meet networks—using systems approaches to infer regulatory networks in plants. *Curr. Opin. Plant Biol.*, **13**, 126–131.

Moreno-Risueno, M.Á., González, N., Díaz, I., Parcy, F., Carbonero, P. and Vicente-Carbajosa, J. (2008) FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and *Arabidopsis*. *Plant J.*, **53**, 882–894.

Muller, M. and Knudsen, S. (1993) The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. *Plant J.*, **4**, 343–355.

Narula, K., Datta, A., Chakraborty, N. and Chakraborty, S. (2013) Comparative analyses of nuclear proteome: extending its function. *Front. Plant Sci.*, **4**, 100.

Nicolas, M.E., Gleadow, R.M. and Dalling, M.J. (1984) Effects of drought and high temperature on grain growth in wheat. *Aust. J. Plant Physiol.*, **11**, 553.

Noctor, G., Gomez, L., Vanacker, H. and Foyer, C.H. (2002) Interactions between biosynthesis, compartmentation and transport in the control of glutathione homeostasis and signalling. *J. Exp. Bot.*, **53**, 1283–1304.

Nuttall, J.G., Leary, G.J.O., Panozzo, J.F., Walker, C.K., Barlow, K.M. and Fitzgerald, G.J. (2016) Models of grain quality in wheat—A review. *F. Crop. Res.*

Oliver, S. (2000) Guilt-by-association goes global. *Nature*, **403**, 601–603.

Osborne, T.B. (1907) The proteins of the wheat kernel. *Carnegie Inst. Washingt.*, **84**, 1–119.

Osborne, T.B. (1909) The vegetable proteins. *Longmans, Green Co.*, **1**, 1–125.

Palta, J. and Fillery, I. (1995) N application increases pre-anthesis contribution of dry matter to grain yield in wheat grown on a duplex soil. *Aust. J. Agric. Res.*, **46**, 519–531.

Pandey, A., Chakraborty, S., Datta, A. and Chakraborty, N. (2008) Proteomics approach to identify dehydration responsive nuclear proteins from chickpea (*Cicer arietinum* L.). *Mol. Cell. Proteomics*, **7**, 88–107.

Payne, P.I., Holt, L.M., Jackson, E. a., Law, C.N. and Damania, a. B. (1984) Wheat storage proteins: their genetics and their potential for manipulation by plant breeding. *Philos. Trans. R. Soc. B Biol. Sci.*, **304**, 359–371.

Payne, P.I., Law, C.N. and Mudd, E.E. (1980) Control by homoeologous group 1 chromosomes of the high-molecular-weight subunits of glutenin, a major protein of wheat endosperm. *Theor. Appl. Genet.*, **58**, 113–120.

Pendle, A.F., Clark, G.P., Boon, R., et al. (2005) Proteomic analysis of the *Arabidopsis* nucleolus suggests novel nucleolar functions. , **16**, 260–269.

Petrovská, B., Jeřábková, H., Chamrád, I., Vrána, J., Lenobel, R., Uřinovská, J., Sebela, M. and Doležel, J. (2014) Proteomic analysis of barley cell nuclei purified by flow sorting. *Cytogenet. Genome Res.*, **143**, 78–86.

Petrovská, B., Šebela, M. and Doležel, J. (2015) Inside a plant nucleus: discovering the proteins. *J. Exp. Bot.*, **66**, 1627–1640.

Pomeranz, Y. (1988) Chemical composition of kernel structures. *Wheat Chem. Technol.*, **2**, 97–158.

Provero, P. (2002) *Gene networks from DNA microarray data: centrality and lethality*, Turin.

Rahman, S., Shewry, P.R., Forde, B.G., Kreis, M. and Miflin, B.J. (1983) Nutritional control of storage protein synthesis in developing grain of barley (*Hordeum vulgare* L.). *Planta*, **159**, 366–372.

Randall, P., Spencer, K. and Freney, J. (1981) Sulfur and nitrogen fertilizer effects on wheat. I. Concentrations of sulfur and nitrogen and the nitrogen to sulfur ratio in grain, in relation to the yield response. *Aust. J. Agric. Res.*, **32**, 203.

Ravel, C., Fiquet, S., Boudet, J., Dardevet, M., Vincent, J., Merlino, M., Michard, R. and Martre, P. (2014) Conserved cis-regulatory modules in promoters of genes encoding wheat high-molecular-weight glutenin subunits. *Front. Plant Sci.*, **5**, 621.

Remans, T., Nacry, P., Pervent, M., Filleur, S., Diatloff, E., Mounier, E., Tillard, P., Forde, B.G. and Gojon, A. (2006) The Arabidopsis NRT1.1 transporter participates in the signaling pathway triggering root colonization of nitrate-rich patches. *Proc. Natl. Acad. Sci.*, **103**, 19206–19211.

Repetto, O., Rogniaux, H., Firnhaber, C., Zuber, H., Küster, H., Larré, C., Thompson, R. and Gallardo, K. (2008) Exploring the nuclear proteome of *Medicago truncatula* at the switch towards seed filling. *Plant J.*, **56**, 398–410.

Repetto, O., Rogniaux, H., Larré, C., Thompson, R. and Gallardo, K. (2012) The seed nuclear proteome. *Front. Plant Sci.*, **3**, 289.

Reussi, N., Echeverría, H. and Rozas, H.S. (2011) Diagnosing sulfur deficiency in spring red wheat: plant analysis. *J. Plant Nutr.*, **34**, 573–589.

Reuveny, Z., Dougall, D.K. and Trinity, P.M. (1980) Regulatory coupling of nitrate and sulfate assimilation pathways in cultured tobacco cells. *Proc. Natl. Acad. Sci. U. S. A.*, **77**, 6670–6672.

Rubio-Somoza, I., Martinez, M., Abraham, Z., Diaz, I. and Carbonero, P. (2006) Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. *Plant J.*, **47**, 269–281.

Rubio-Somoza, I., Martinez, M., Diaz, I. and Carbonero, P. (2006) HvMCB1, a R1MYB transcription factor from barley with antagonistic regulatory functions during seed development and germination. *Plant J.*, **45**, 17–30.

Sabelli, P.A. and Larkins, B.A. (2009) The development of endosperm in grasses. *Plant Physiol.*, **149**, 14–26.

Scherer, H.W. (2001) Sulphur in crop production. *Eur. J. Agron.*, **14**, 81–111.

Schnyder, H. and Baum, U. (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.

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M. and Ligterink, W. (2016) Learning from co-expression networks: possibilities and challenges. *Front. Plant Sci.*, **7**, 444.

Shewry, P.R. (2009) Wheat. *J. Exp. Bot.*, **60**, 1537–53.

Shewry, P.R. and Halford, N.G. (2003) Genetics of wheat gluten proteins. *Adv. Genet.*, **49**, 111–184.

Shewry, P.R. and Hey, S.J. (2015) The contribution of wheat to human diet and health. *Food Energy Secur.*, **4**, 178–202.

Shewry, P.R., Mitchell, R.A.C., Tosi, P., et al. (2012) An integrated study of grain development of wheat (cv. Hereward). *J. Cereal Sci.*, **56**, 21–30.

Shewry, P.R., Tatham, A.S., Forde, J., Kreis, M. and Miflin, B.J. (1986) The classification and nomenclature of wheat gluten proteins: A reassessment. *J. Cereal Sci.*, **4**, 97–106.

Shewry, P.R., Tatham, A.S. and Halford, N.G. (2001) Nutritional control of storage protein synthesis in developing grain of wheat and barley. *Plant Growth Regul.*, **34**, 105–111.

Shewry, P.R., Tatham, A.S. and Lazzeri, P. (1997) Biotechnology of wheat quality. *J. Sci. Food Agric.*, **73**, 397–406.

Silva, A.T., Ribone, P.A., Chan, R.L., Ligterink, W. and Hilhorst, H.W.M. (2016) A predictive coexpression network identifies novel genes controlling the seed-to-seedling phase transition in *Arabidopsis thaliana*. *Plant Physiol.*, **170**, 2218–2231.

Simmonds, N.W. (1995) The relation between yield and protein in cereal grain. *J. Sci. Food Agric.*, **67**, 309–315.

Skylas, D.J., Cordwell, S.J., Hains, P.G., et al. (2002) Heat shock of wheat during grain filling: proteins associated with heat-tolerance. *J. Cereal Sci.*, **35**, 175–188.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Subba, P., Kumar, R., Gayali, S., Shekhar, S., Parveen, S., Pandey, A., Datta, A., Chakraborty, S. and Chakraborty, N. (2013) Characterisation of the nuclear proteome of a dehydration-sensitive cultivar of chickpea and comparative proteomic analysis with a tolerant cultivar. *Proteomics*, **13**, 1973–92.

Surget, A. and Barron, C. (2005) Histologie du grain de bté. *Ind. des céréales*, **145**, 3–7.

Tamura, K. and Hara-Nishimura, I. (2014) Functional insights of nucleocytoplasmic transport in plants. *Front. Plant Sci.*, **5**, 118.

Tamura, K. and Hara-Nishimura, I. (2013) The molecular architecture of the plant nuclear pore complex. *J. Exp. Bot.*, **64**, 823–832.

Tan, F., Li, G., Chitteti, B.R. and Peng, Z. (2007) Proteome and phosphoproteome analysis of chromatin associated proteins in rice (*Oryza sativa*). *Proteomics*, **7**, 4511–27.

Taylor-Teeples, M., Lin, L., Lucas, M. de, et al. (2014) An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature*, **517**, 571–575.

Triboï, E., Martre, P. and Triboï-Blondel, A.-M. (2003) Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. *J. Exp. Bot.*, **54**, 1731–1742.

Urano, K., Kurihara, Y., Seki, M. and Shinozaki, K. (2010) "Omics" analyses of regulatory networks in plant abiotic stress responses. *Curr. Opin. Plant Biol.*, **13**, 132–138.

Usadel, B., Obayashi, T., Mutwil, M., et al. (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant. Cell Environ.*, **32**, 1633–1651.

USDA (2016) Grain: World Markets and Trade TMT.

Vensel, W.H., Tanaka, C.K., Cai, N., Wong, J.H., Buchanan, B.B. and Hurkman, W.J. (2005) Developmental changes in the metabolic protein profiles of wheat endosperm. *Proteomics*, **5**, 1594–611.

Verdier, J. and Thompson, R.D. (2008) Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.*, **49**, 1263–1271.

Vincent, J., Martre, P., Gouriou, B., Ravel, C., Dai, Z., Petit, J.-M. and Pailloux, M. (2015) RulNet: A web-oriented platform for regulatory network inference, application to wheat – omics data. *PLoS One*, **10**, e0127127.

Wieser, H., Gutser, R. and Tucher, S. von (2004) Influence of sulphur fertilisation on quantities and proportions of gluten protein types in wheat flour. *J. Cereal Sci.*, **40**, 239–244.

Wieser, H. and Seilmeier, W. (1998) The influence of nitrogen fertilisation on quantities and proportions of different protein types in wheat flour. *J. Sci. Food Agric.*, **76**, 49–55.

Wrigley, C., Cros, D., Archer, M., Downie, P. and Roxburgh, C. (1980) The sulfur content of wheat endosperm proteins and its relevance to grain quality. *Aust. J. Plant Physiol.*, **7**, 755–766.

Xi, D.-M. and Zheng, C.-C. (2011) Transcriptional regulation of seed storage protein genes in Arabidopsis and cereals. *Seed Sci. Res.*, **21**, 247–254.

Yamaguchi, Y., Nakamura, T., Harada, E., Koizumi, N. and Sano, H. (1999) Differential accumulation of transcripts encoding sulfur assimilation enzymes upon sulfur and/or nitrogen deprivation in *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.*, **63**, 762–766.

Young, T.E. and Gallie, D.R. (2000) Programmed cell death during animal development. *Plant Mol. Biol.*, **44**, 283–301.

Zhao, F., Evans, E.J., Bilsborrow, P.E. and Syers, J.K. (1993) Influence of sulphur and nitrogen on seed yield and quality of low glucosinolate oilseed rape (*Brassica napus* L). *J. Sci. Food Agric.*, **63**, 29–37.

Zhao, F., Hawkesford, M. and McGrath, S. (1999) Sulphur assimilation and effects on yield and quality of wheat. *J. Cereal Sci.*, **30**, 1–17.

Zörb, C., Grover, C., Steinfurth, D. and Hermann Mühling, K. (2010) Quantitative proteome analysis of wheat gluten as influenced by N and S nutrition. *Plant Soil*, **327**, 225–234.

CHAPITRE 2 :
Validation de la méthode d'extraction des protéines nucléaires à partir de grains de blé

1. INTRODUCTION

Pour comprendre les mécanismes impliqués dans la régulation du développement du grain et l'accumulation de ses réserves, l'étude des protéines nucléaires est apparue nécessaire. Pour cela, nous avons mis au point chez le blé une méthode nous permettant d'accéder à ce type de protéines généralement sous-représentées dans les analyses de protéomique. Parmi les études menées chez les plantes, la méthode la plus couramment utilisée repose sur la centrifugation de l'extrait cellulaire dans un gradient de densité pour purifier les noyaux. Dans le grain, la principale difficulté pour la purification d'organites réside dans la présence en abondance des composés de réserves. Ceci pourrait expliquer pourquoi peu d'études se sont intéressées au protéome nucléaire du grain (Repetto *et al.*, 2012).

La procédure d'obtention d'extraits de protéines nucléaires utilisée au cours de cette thèse a été mise au point par Emmanuelle Bancel avant mon arrivée au sein de l'équipe BIG. Cette méthode est composée de deux étapes principales : 1/ la purification de noyaux à partir de grains de blé, 2/ l'extraction des protéines des noyaux.

Pour décrire brièvement cette procédure, les grains sont tout d'abord broyés en milieu liquide dans un tampon d'extraction, puis l'homogénat est filtré afin d'éliminer les débris cellulaires. Les membranes cellulaires sont ensuite lysées par l'ajout d'un détergent, le Triton X-100, et les noyaux sont purifiés sur un gradient discontinu 30%/80% de Percoll, où ils s'accumulent à l'interface des deux solutions de densité différentes. Les protéines sont extraites de ces noyaux par l'utilisation du TriReagent, qui est une solution monophasique de phénol et de guanidine isothiocyanate, permettant la séparation de l'extrait nucléaire en trois phases : la phase aqueuse contenant les ARN, l'interphase contenant l'ADN, et la phase organique qui contient les protéines nucléaires. Les protéines nucléaires extraites peuvent alors être analysées en spectrométrie de masse, avec ou sans séparation préalable sur gel mono- ou bi-dimensionnel.

Afin de valider la méthode, plusieurs expérimentations auxquelles j'ai participé ont été entreprises. Ainsi, des colorations de noyaux au Hoechst®, des Western blot utilisant des anticorps dirigés contre des protéines marqueurs des différents compartiments subcellulaires, puis une identification des protéines par spectrométrie de masse ont été réalisés. Ces étapes de validation ont été réalisées pour des grains de *T. aestivum* et *T.monococcum*, les deux espèces de blé utilisées au cours de la thèse, récoltés à deux temps thermiques, 150 et 250°Cj après floraison. Pour l'identification des protéines, nous avons fait le choix de les analyser en spectrométrie de masse à l'issue de leur extraction, sans séparation sur gel.

Références :

Repetto, O., Rogniaux, H., Larré, C., Thompson, R. and Gallardo, K. (2012) The seed nuclear proteome. *Front. Plant Sci.*, **3**, 289.

2. ARTICLE 1 : Proteomic approach to identify nuclear proteins in wheat grain

Emmanuelle Bancel[1,2], Titouan Bonnot[1,2], Marlène Davanture[3], Gérard Branlard[1,2], Michel Zivy[3], and Pierre Martre[1,2]

[1]INRA, UMR1095 Génétique, Diversité et Ecophysiologie des Céréales, 5 chemin de Beaulieu, F-63 039 Clermont-Ferrand, France
[2]Université Blaise Pascal, UMR1095 Génétique, Diversité et Ecophysiologie des Céréales, Avenue des Landais, F-63 170 Aubière, France
[3]CNRS, PAPPSO, UMR 0320/8120 Génétique Quantitative et Évolution - Le Moulon, F-91190 Gif-sur-Yvette, France

Journal of proteome research

Technical Note

pubs.acs.org/jpr

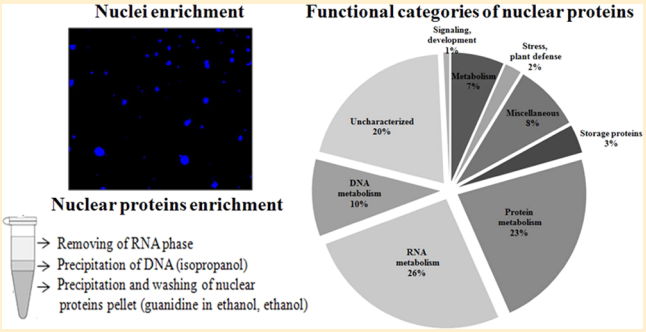
Proteomic Approach to Identify Nuclear Proteins in Wheat Grain

Emmanuelle Bancel,[*,†,‡,∥] Titouan Bonnot,[†,‡,∥] Marlène Davanture,[§] Gérard Branlard,[†,‡] Michel Zivy,[§] and Pierre Martre[†,‡,⊥]

[†]INRA, UMR1095 Genetics, Diversity and Ecophysiology of Cereals, 5 chemin de Beaulieu, F-63 039 Clermont-Ferrand, France
[‡]Blaise Pascal University, UMR1095 Genetics, Diversity and Ecophysiology of Cereals, Avenue des Landais, F-63 170 Aubière, France
[§]CNRS, PAPPSO, UMR 0320/8120 Génétique Quantitative et Évolution - Le Moulon, F-91190 Gif-sur-Yvette, France

Supporting Information

ABSTRACT: The nuclear proteome of the grain of the two cultivated wheat species *Triticum aestivum* (hexaploid wheat; genomes A, B, and D) and *T. monococcum* (diploid wheat; genome A) was analyzed in two early stages of development using shotgun-based proteomics. A procedure was optimized to purify nuclei, and an improved protein sample preparation was developed to efficiently remove nonprotein substances (starch and nucleic acids). A total of 797 proteins corresponding to 528 unique proteins were identified, 36% of which were classified in functional groups related to DNA and RNA metabolism. A large number (107 proteins) of unknown functions and hypothetical proteins were also found. Some identified proteins may be multifunctional and may present multiple localizations. On the basis of the MS/MS analysis, 368 proteins were present in the two species, and in two stages of development, some qualitative differences between species and stages of development were also found. All of these data illustrate the dynamic function of the grain nucleus in the early stages of development.



KEYWORDS: *cereal, grain development, bread wheat (Triticum aestivum), einkorn wheat (Triticum monococcum), LC−MS/MS, nuclear proteome*

INTRODUCTION

Over the past 10 years, several studies have focused on cereal grain proteome. For wheat, both global[1−3] and targeted studies[4−7] were developed; however, little is known about subcellular proteome like nuclear proteome. Nuclear proteins are predicted to comprise ∼20% of the total cellular proteins in yeast and animals, but their amount in plants is not yet known.[8] Several nuclear proteomics investigations have been performed in tissues and organs other than grains like leaves,[9] cell suspensions[10,11] or seedlings.[12] Some studies have concerned whole seeds for *Medicago truncatula*[13] or endosperm for *Oryza sativa*[14] or *Zea mays*.[15] Nuclear proteins are often underrepresented in proteomic studies due to the frequently low abundance of proteins involved in regulatory processes.[16,17]

Knowledge of the changes in nuclear proteome that occur during grain development is needed to improve our understanding of the specific mechanisms and involved in endosperm cell regulation. Indeed to understand the mechanism associated with grain development, it is of prime importance to know which nuclear actors are present and to develop a proteomic approach allowing qualitative and quantitative studies; however, the nuclear proteome of grains and of wheat, in particular, has been far less studied. Two major difficulties may explain why cereal grain nuclear proteome was less studied: (i) in the endosperm, nuclei are entrapped in protein and starch matrix, making difficult the isolation of pure nuclei, and (ii) numerous compounds, particularly carbon (starch or oil) and nitrogen (storage proteins) storage components and nucleic acids, may interfere during protein extraction and proteomic analysis. Therefore, optimizations are necessary to obtain high-purity nuclei in all developmental stages, which differ in the number of contaminants (e.g., protein bodies, starch granules).

The purpose of the present note is first to report on an extraction protocol allowing nuclear proteome analysis to be performed on developing cereal grains, with the specific objective of obtaining nuclei in sufficient yield and purity for proteomic approach. Because our aim is to use this protocol in further studies to analyze the response of the nuclear proteome for a wide range of genetic and environmental contexts, the method has to produce a low number of protein fractions and to be adapted with shotgun proteomics. The second objective is to provide a qualitative overview of the nuclear proteome revealed through gel-free mass spectrometry analysis during the early phases of endosperm cells division and cells differentiation. These early phases of grain development determine the potential final size of the grain,[18] and they are particularly sensitive to the growing temperature[19] and to soil water

Received: May 21, 2015
Published: July 31, 2015

DOI: 10.1021/acs.jproteome.5b00446
J. Proteome Res. 2015, 14, 4432−4439

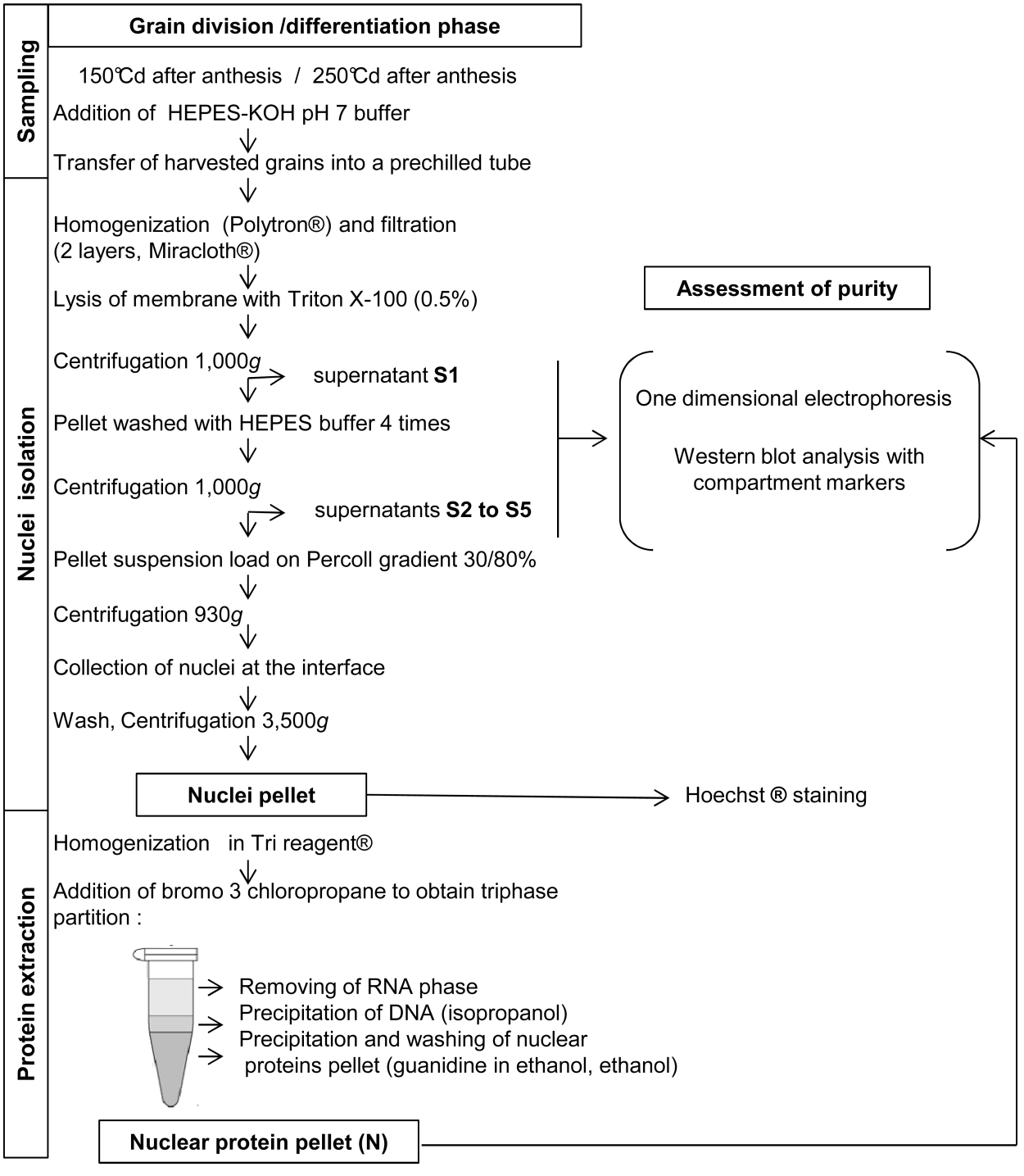


Figure 1. Workflow with the main steps used to purify nuclei and extract nuclear proteins from wheat grain.

deficit.[20] The methodology was validated for the cultivated hexaploid wheat species *Triticum aestivum* (genomes A, B, and D) and its domesticated diploid progenitor *T. monococcum* (genome A).

METHODS

Plant Material and Tissue Sampling

The bread wheat (*Triticum aestivum*) cultivar Récital and the einkorn wheat (*Triticum monococcum*) accession FR35821 - K48993 were cultivated in an environmentally controlled growth chamber at INRA, Clermont-Ferrand, France. Plants were grown under natural soil conditions, fertilized, watered as usual, and protected against fungi. Ears were tagged at the date of anthesis, and only grains from the middle of the tagged ears were harvested. Air temperature close to the ears was recorded, and daily mean air temperature was calculated, and the sum of daily mean temperatures was used to follow the grain development in thermal time (degree-days, °Cd). Wheat grains were harvested at two different thermal times after anthesis corresponding to the endosperm cells division (150 °Cd after anthesis) and cell differentiation stages (250 °Cd after anthesis). For each developmental stage, grains from the middle third parts of the ears were quickly harvested, weighed,

DOI: 10.1021/acs.jproteome.5b00446
J. Proteome Res. 2015, 14, 4432−4439

frozen in liquid nitrogen, and then stored at −80 °C until use. To validate the technical and methodological approach and for protein quantification, we used at least four biological replicates.

Isolation of Grain Nuclei

A method (Figure 1) was adapted from ref 21. All steps were performed at 4 °C. Grains (2 g) were put into ice-cold round-bottomed centrifuge tubes immersed in 6 mL of extraction buffer (20 mM Hepes-KOH (pH 7.0), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.5 mM PMSF, and 0.1% (v/v) phosphatase inhibitors cocktail (Sigma-Aldrich) added freshly). Tissues were disrupted with a polytron homogenizer (Kinematica POLY-TRON PT 10), and the resultant homogenate was filtered through two layers of Miracloth (Calbiochem, pore size 25 μm) to remove cell debris. Membranes were then lyzed by adding Triton X-100 to the homogenate (adjusted to 12 mL) at a final concentration of 0.5% (v/v). After centrifugation (1000*g* for 5 min), the resultant supernatant adjusted to 12 mL was kept (S1), and the pellet was washed four times with 1 mL of extraction buffer, followed by centrifugation (1000*g* for 5 min; S2−S5) to eliminate contaminants (mainly starch and storage proteins) and obtain semipure nuclei. These nuclei were thoroughly suspended in 1 mL of extraction buffer.

The resulting suspension was layered in a 16 mL tube on top of a Percoll gradient consisting of 10 mL of a 30% (v/v) Percoll solution and 2.5 mL of a 80% (v/v) Percoll solution. After centrifugation (930*g* for 30 min), nuclei were collected at the interface of the 30 and 80% Percoll layers, washed twice with extraction buffer, and recovered as a pellet after centrifugation (3500*g* for 5 min). The nuclei pellet was suspended in phosphate-buffered saline, and an aliquot was stained in a 0.1 μg mL^{-1} Hoechst fluorescent dye solution and observed under fluorescence microscopy (Zeiss Axioplan) to evaluate the nuclear enrichment and integrity.

Extraction of Nuclear Proteins

The nuclei-containing pellets were resuspended in TRI Reagent (Sigma-Aldrich) according to the manufacturer's instruction. This product, a mixture of guanidine thiocyanate and phenol in a monophase solution, effectively dissolves DNA, RNA, and protein. After adding 1-bromo 3-chloropropane and centrifuging, the mixture separates into three phases: an aqueous phase containing the RNA, the interphase containing DNA, and an organic phase containing proteins. Each component can then be isolated after separating the phases (Figure 1). In brief, nuclei pellet was resuspended in 500 μL of TRI Reagent and incubated for 5 min at room temperature. Then, 50 μL of bromo-3-chloropropane was added, and the mixture was vortexed for 15 s and incubated for 15 min at room temperature. After centrifugation (12 000*g* for 15 min at 4 °C), the upper phase corresponding to RNA was removed. To precipitate DNA, 150 μL of 95% (v/v) ethanol was added. After mixing by inversion and incubation for 3 min at room temperature, the mixture was centrifuged (2000*g* for 5 min at 4 °C) and the supernatant was collected. Finally, proteins were precipitated by adding 750 μL of 2-propanol, incubating for 10 min at room temperature and centrifuging (12 000*g* for 10 min at 4 °C). Supernatant was removed and protein pellet was washed three times in 1 mL of 0.3 M guanidine hydrochloride/95% ethanol solution by stirring for 20 min, followed by centrifugation (7500*g* for 5 min at 4 °C). The protein pellet was then washed with 1 mL of 95% (v/v) ethanol. The final protein pellet was dried and then stored at −20 °C.

SDS-PAGE Analysis

A fraction (1 mL) of each supernatant (S1−S5) was precipitated with TCA (w/v) 10% in cold acetone and proteins present in the pellet were solubilized in SDS-PAGE loading buffer (45 mM Tris-HCl, pH 6.8, 50 mM DTT, 1% SDS, 10% glycerol, 0.001% bromophenol blue) as well as nuclear proteins pellet to check the different supernatants during the washing steps. Proteins were loaded at constant volume and separated by SDS-PAGE in a 12.5% polyacrylamide gel. Electrophoresis was performed at 70 V for 0.3 h and then at 200 V for 1 h using a Mini-Protean II gel electrophoresis system (Bio-Rad). The gels were stained with 0.2% Coomassie brilliant blue R250 (Sigma-Aldrich) for 1 h at room temperature and subsequently destained in 10% (v/v) acetic acid and 30% (v/v) ethanol. Protein concentration in nuclear extracts was determined by Bradford assay with bovine serum albumin as standard.[22]

Assessment of the Purity of the Nuclei Fractions

For immunoblot analysis, protein samples corresponding to nuclear proteins and to proteins from supernatants collected during nuclei enrichment were separated on a 12.5% SDS−PAGE gel and then transferred onto a nitrocellulose membrane (Hybond, ECL, GE Healthcare) using a Hoefer TE77 semidry transfer blotter. The blotted membrane was incubated for 1 h at room temperature in blocking buffer consisting of 10 mM Tris−HCl (pH 7.6), 150 mM NaCl, 0.01% (v/v) Tween 20, and 5% (w/v) skim milk. After blocking, the membrane was incubated 1 h at room temperature and then overnight at 4 °C with a 1:1000 dilution of antihistone H3 (Abcam) antibody, antivacuolar ATPase (Agrisera AB) antibody, antiphotosystem II reaction center protein D1 (Agrisera AB), or with a 1:2500 dilution of anti-UDP-glucose pyrophosphorylase (Agrisera AB) and anti-H protein of glycine decarboxylase complex (GDC; Agrisera AB). Antirabbit IgG conjugated with horseradish peroxidase (GE Healthcare) was used as the secondary antibody (1:100 000). After 1 h of incubation with the secondary antibody, signals were detected using an ECL plus Western blotting detection kit (GE Healthcare) following the manufacturer's protocol.

Trypsin Digestion of Nuclear Proteins

For digestion, a protein sample for each species (*Triticum aestivum* and *Triticum monococcum*) in each developmental stage (150 and 250 °Cd after anthesis) was analyzed. Precipitated proteins were suspended in 30 μL of buffer A (0.1% ZALS I, 6 M urea, 2 M thiourea, 10m M DTT, 30 mM Tris-HCl pH 8.8, 5 mM NH_4HCO_3). Protein concentration was measured using the 2-D Quant Kit (GE Healthcare, Piscataway, NJ) with BSA as a standard. All samples were adjusted to a single concentration of 5 μg/μL by adding buffer A. Then, 40 μg of protein of each sample was incubated with iodoacetamide (40 mM final concentration) in the dark for 1 h. They were then diluted 8-fold with 50 mM NH_4HCO_3. Samples were digested in-solution overnight at 37 °C by adding 800 ng of trypsin in 50 mM NH_4HCO_3 (trypsin/protein ratio 1/50). Trypsic digestion was stopped by the addition of TFA 10%.

LC−MS/MS Analysis

Online liquid chromatography was performed on a NanoLC Ultra system (Eksigent). A 1 μg sample was loaded at 7.5 μL min^{-1} on a precolumn (stationary phase: C18, particles of 5 μm; column: 100 μm i.d., 2 cm length; NanoSeparations) and desalted with 0.1% (v/v) formic acid in 2% (v/v) ACN. After 3

DOI: 10.1021/acs.jproteome.5b00446
J. Proteome Res. 2015, 14, 4432−4439

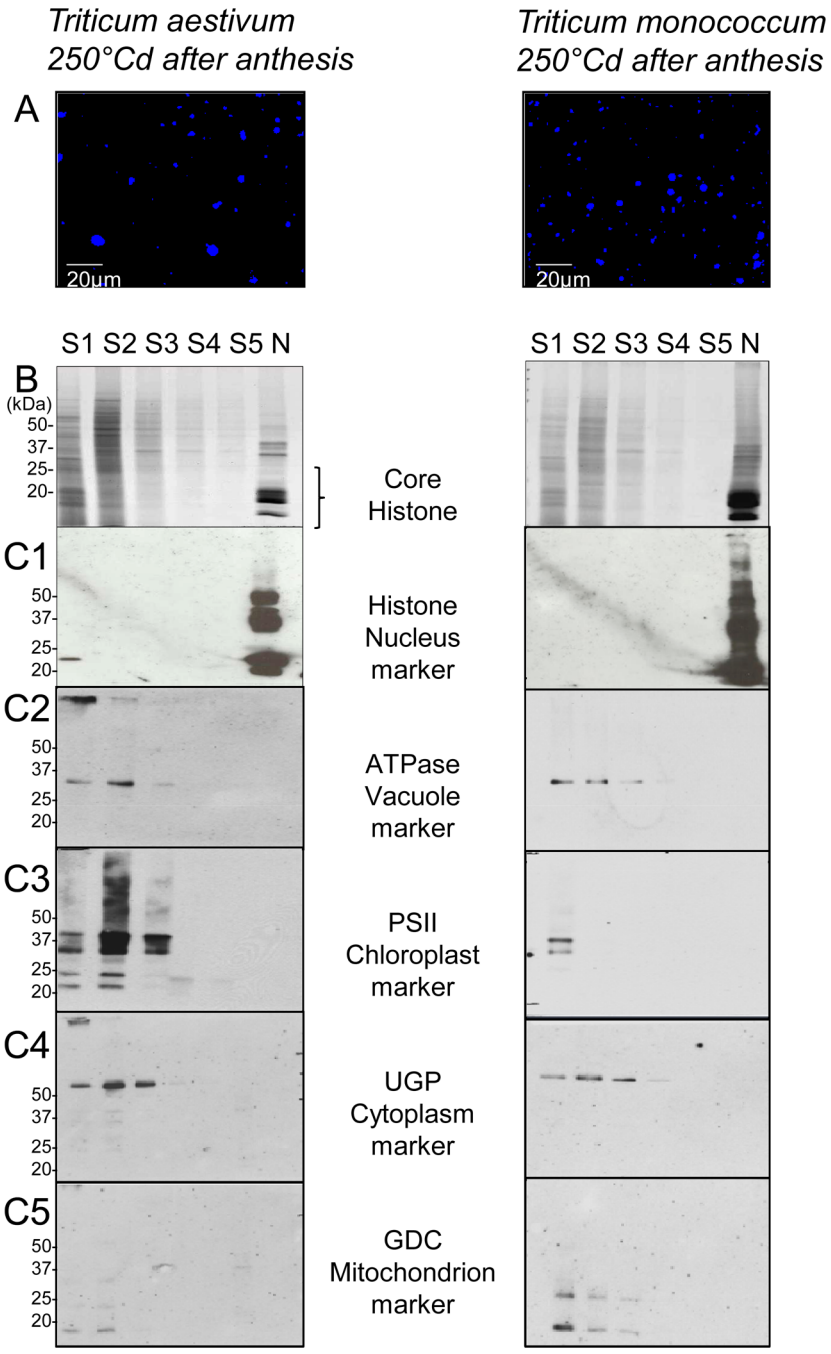


Figure 2. Assessment of the purity of nuclei preparation and nuclear protein extract (N) in *Triticum aestivum* (left) and *Triticum monococcum* (right). (A) Purified nuclei fraction was stained with Hoechst and visualized by fluorescence microscopy. (B) Analytical 1-D electrophoresis profile (12.5% SDS-PAGE, Coomassie Blue staining) of protein content in different stages of nuclei enrichment (S1−S5, N). (C) (1) Immunoblot analysis with antihistone H3, (2) antivacuolar ATPase, (3) antiphotosystem II reaction center protein D1, (4) anti-UDP-glucose pyrophosphorylase, and (5) anti-H protein of glycine decarboxylase complex (GDC) antibodies.

min, the precolumn was connected to a separating C18 column (stationary phase BIOSPHERE C18, particles of 3 μm; column 75 μm i.d., 150 mm length; NanoSeparations). Buffers were 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Peptide separation was achieved using a linear gradient from 5 to 30% B for 28 min at 300 nL min^{-1}. Including the regeneration step at 100% B and the equilibration step at 100% A, one run took 45 min. Eluted peptides were analyzed with a Q-Exactive mass spectrometer (Thermo Electron, Courtaboeuf, France) using a nanoelectrospray interface. Ionization was performed with a 1.3 kV spray voltage applied to an uncoated capillary probe (10 μ i.d.; New Objective, Woburn, MA). Xcalibur 2.2 SP1 interface was used to monitor data-dependent acquisition of peptide ions. This included a full MS scan covering the 300 to 1400 mass-to-charge ratio (m/z) with a resolution of 70 000 and a MS/MS step (normalized collision energy: 30%; resolution: 17 500). The MS/MS step was reiterated for the eight major ions detected during the full MS scan. The dynamic exclusion was set to 45 s.

DOI: 10.1021/acs.jproteome.5b00446
J. Proteome Res. 2015, 14, 4432−4439

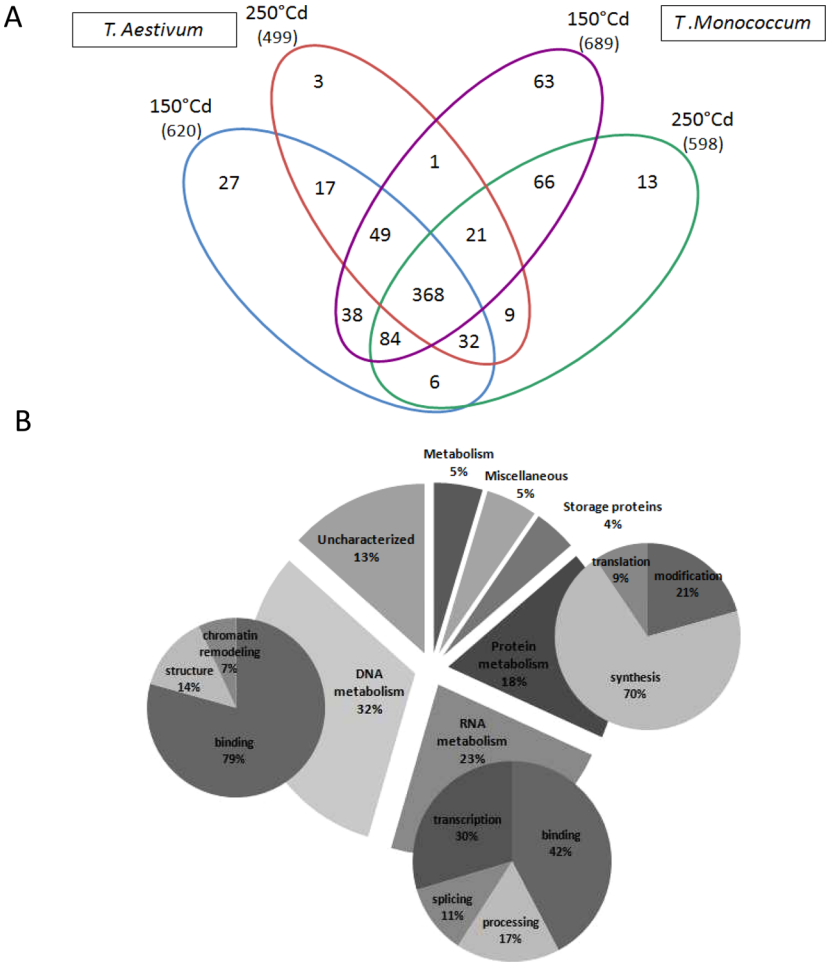


Figure 3. Graphical representation of the functional categories of proteins found in wheat grain nuclei-enriched fraction. (A) Venn diagrams analysis of the 797 proteins identified for the two species (*Triticum aestivum* and *Triticum monococcum*) in the two developmental stages (150 and 250 °Cd after anthesis). (B) Graphical representation of the functional categories of the 368 commun proteins for the two species (*Triticum aestivum* and *Triticum monococcum*) in the two developmental stages (150 and 250 °Cd after anthesis) according to Gene ontology. The percentage of proteins in each functional class is shown.

Identification of Nuclear Proteins

Protein identification was performed by searching peptides against the Uniprot protein database version 2014_07 limited to the *Triticum* genus using X!Tandem (version: Sledge-hammer, 2013.09.01.1, http://www.thegpm.org/TANDEM). Enzymatic cleavage parameters were set as trypsin digestion with one possible miss cleavage. Cys carboxyamidomethylation was set as static modification, whereas Met oxidation, Nter deamidation, and Nter acetylation were set as variable modifications. Precursor mass tolerance was 10 ppm and fragment mass tolerance was 0.02. Identified proteins were filtered and grouped using X!TandemPipeline v3.3.1 (http://pappso.inra.fr/bioinfo/xtandempipeline/). Peptide and protein *e*-value cut offs were set to 0.01 and 10^{-4}, respectively, with at least two peptides per protein. Proteins with at least one peptide in common and similar function were defined as unique proteins. The presence of several proteins identified by the same sets of peptides in a group showed the redundancy due to the approach itself giving a list of all proteins that might be present in the sample.

Functional Classification of Nuclear Proteins and Subcellular Localization

Functional classification was established on the basis of gene ontology (GO) information rules[23] provided by Uniprot. Subcellular localization of identified proteins was predicted by in silico analysis using Multiloc2,[24] NucPred,[25] WolfPSort,[26] Y loc,[27] and LocTree[28] programs. The top two hits were considered for Multiloc and Y loc. Some proteins were also detected and localized from previous studies in the literature.

RESULTS AND DISCUSSION

Purity of Nuclei Preparation

The development of wheat grain involves three distinct phases: cell division and differentiation, grain filling, and desiccation/maturation. At 110 °Cd (6 days) after anthesis, nuclei are present in the central region of the endosperm that was occupied by the central vacuole because of cellularization of peripheral endosperm. All major cell types are differentiated around 250 °Cd (13 days) after anthesis.[29] The nuclei were isolated from grains in two stages corresponding to this first phase of development. Different methods were used for assessing organelle intactness and purity. The Hoechst fluorescence image obtained assesses the integrity of nuclei

DOI: 10.1021/acs.jproteome.5b00446
J. Proteome Res. 2015, 14, 4432−4439

(Figure 2A). The nuclei had average diameter of approximately 5−10 μm.

The purity of isolated organelles can be compromised by copurifying subcellular fractions. Proteins from supernatants (S1−S5) collected during nuclei purification were solubilized and analyzed by SDS-PAGE. We extracted proteins from the nuclei-enriched fraction using the TRI Reagent extraction method and in loading this nuclear extract (N) on SDS PAGE. The average final yield of nuclear proteins extraction for *T. aestivum* was 160 and 110 $\mu g\ g^{-1}$ fresh mater (FM) at 150 and 250 °Cd after anthesis, respectively, and for *T. monococcum* was 120 and 150 $\mu g\ g^{-1}$ FM at 150 and 250 °Cd after anthesis, respectively.

The combination of microscopy and biochemical methods was used to assess the purity of our nuclei and nuclear proteins preparations. Coomassie blue staining revealed the absence of significant amounts of proteins in the last supernatant (S5) and a typical nuclear protein pattern with intense bands corresponding to histones (Figure 2C1). Therefore, we tested the presence of possible contamination from other subcellular compartments by assaying the presence of the representative marker enzymes for those compartments (i.e., mitochondrion, chloroplast, thylakoids, and cytoplasm) by Western immunoanalyses. The Western blot analyses performed with the different antibodies showed an enrichment in low-molecular-weight proteins in nuclear extracts corresponding to histone variants (Figure 2C1) and the absence of vacuolar ATPase, photosystem II reaction center protein D1, UDP-glucose pyrophosphorylase, and H protein of glycine decarboxylase complex (GDC) in these extracts, confirming no detectable level of cross contamination (Figure 2C2−C5).

Our protocol as compared with the initial one included additional washing steps in which contaminants are removed from the nuclei-enriched pellet without important loss of nuclear proteins, as shown on Western blots antihistone (Figure 2C1), and also an improved method for qualitative and quantitative extraction of nuclear proteins.

Protein Identification through Mass Spectrometry

Shotgun proteomics, which is the analysis of a complex mixture of peptides using LC−MS/MS without prior separation, was employed in this study to analyze nuclei protein fraction. The databank was Uniprot and the search program used was Xtandem (Release 2013.9.1.0). The proteins were identified with at least two of their peptides matching the database entry with a P value <0.01. By this approach, 797 proteins were identified with high confidence. These proteins were distributed in 528 groups of proteins, which shared at least one common peptide and had the same function. The taxonomy was *Triticum*, so identifications take places in *Triticum aestivum* or in *Triticum urartu*, the diploid progenitor of the bread wheat A-genome (Table S1, Supporting Information).

Unique proteins were classified in nine functional categories according to the GO rules (Figure S1, Supporting Information). The most represented categories were: RNA metabolism (including processing, splicing, transcription, and binding subclasses), protein metabolism (including synthesis (ribosome biogenesis), translation, and modification (including degradation and folding) subclasses), and DNA metabolism (structure, binding, and chromatin remodeling). Twenty percent of the proteins were uncharacterized. Numerous proteins were involved in metabolism and miscellaneous (15%). Some storage proteins (3%) were found and could be considered as contaminants present in small amount in the nuclear extracts. Nevertheless, a nuclear protein belonging to the vicilin superfamily was related to the protection of chromatin structure against desiccation during grain development in *Pisum sativum*,[30] and in *Medicago truncatula* vicilin (7S globulin family) and legumin (11S globulin family), storage proteins were found in grain nuclear extracts.[13] The comparison of identified proteins revealed a core of 368 proteins (46%) that were present in the four samples (Figure 3A). They were principally related to DNA (32%), RNA (23%), and protein (18%) metabolisms (Figure 3B). These data assess the robustness of the approach that enables us to detect a pool of proteins involved in basal nuclear metabolism. Radar plots (Figure S2, Supporting Information) show similar distribution of functional categories between the two stages for the two species. For the two species, an increase in the percentage of proteins involved in DNA metabolism was observed at 250 °Cd anthesis compared with 150 °Cd after anthesis. Interestingly, a higher proportion of proteins involved in RNA metabolism and a higher ratio of uncharacterized proteins was observed in *T. Monococcum* compared with *T. aestivum*, suggesting that these proteins could be expressed at different levels or have different regulatory processes in the two species.

Subcellular Localization

Five different prediction tools were employed to investigate subcellular localization (Table S1, Supporting Information). Eighty four percent of identified unique proteins had a predicted nuclear localization with at least one tool prediction and 65% with at least two tools, which suggest that these proteins are nuclear actors or spend at least some time in the nucleus. Nevertheless, the proteins that were not predicted as having a nuclear localization (16%) cannot be totally excluded due to the limitations of the prediction tools and the fact that the subcellular localization pattern changes as the proteins can be rapidly exported. In eukaryotic cells, up to 35% of proteins have multisubcellular localizations.[31] Proteins may simultaneously locate or move between different cellular compartments, such as transcription factors and signaling pathway transduction factors. Proteins may play different roles in biological processes when they are in different subcellular compartments. For these proteins, single subcellular localization annotation will lose some important information.[31] Sometimes proteins had an atypical nuclear localization, but the fact that they were found in several nuclear proteome studies increases the probability of a nuclear localization. This was the case for some proteins discussed later.

Proteins Having an Atypical Nuclear Localization

Actin-97 belonging to the actin family was identified. Nucleoskeletal proteins such as actin are recognized as having fundamental role in nuclear structure.[32,33] Actin was also found in nuclear proteomes from *Arabidopsis thaliana* leaves[9] and *Cicer arietinum* seedlings.[12]

A 14−3−3 protein was also detected and was already found in *A. thaliana* leaves nuclear proteome.[9] This large family of ∼30 kDa acidic proteins has been implicated in the modulation of distinct biological processes by binding to phosphorylated proteins.[34] Many nuclear proteins are putatively phosphorylated,[35] and some protein kinases can be translocated into nuclei where they phosphorylate transcription factors, which, in turn, modulate gene expression.[36] A protein serine/threonine kinase with potential nuclear localization was also present in our nuclear extracts.

DOI: 10.1021/acs.jproteome.5b00446
J. Proteome Res. 2015, 14, 4432−4439

Five ATP synthases were identified. These ubiquitous membrane enzymes play a key role in biological energy metabolism. Orthologs of mitochondrial ATP synthase have been detected in *A. thaliana* nucleus and other organelles.[37,31]

Several metabolic enzymes have been shown to function as transcriptional regulators. For example, Malate dehydrogenase can have a nucleocytoplasmic localization and be involved in transcription regulation.[38] An enzyme involved in carbohydrate metabolism (glyceraldehyde-3-phosphate dehydrogenase GAPDH) identified in our samples was also found in *Pisum sativum* leaf nuclei.[39] Interestingly, this GAPDH acts as a coactivator to regulate the expression of histone H2B.[40]

The existence of multifaceted roles of glycolytic proteins suggests that links between metabolic sensors and transcription are established directly through enzymes that participate in metabolism.[41] Nothing is known yet about the mechanisms underlying the nuclear/cytoplasmic distribution of metabolic enzymes.

Uncharacterized Proteins

In this study, many proteins (20%) did not have similarity to known proteins and are therefore "'uncharacterized"'. The assigned function can be inappropriate if automated transfer to a distant species is used.[42] No functional annotations based on Blastp similarity searches were given for those proteins. Furthermore, the number of uncharacterized proteins is close to that reported in others nuclear proteome studies with 10−18% of unknown proteins[8] and up to 44% of hypothetical proteins in rice nuclear endosperm.[14] In our study, 70% of these proteins have a predicted nuclear localization with at least two tools (Table S1, Supporting Information).

To explain such number of unidentified proteins, we may hypothesis that even when employing databases of closely related species a large number of viable tandem mass spectra of peptides might not be assigned accurately to a protein, as a single amino acid mutation may significantly alter the peptide mass and the resulting fragmentation pattern.

CONCLUSIONS

In this study, we report the first analysis of the nuclear proteome in developing grains of two wheat species. The purity of the isolated nuclei together with the improved nuclear proteins extraction allows us to perform specific proteomics analysis by Western blot and shotgun approach (this work) but also by 2D electrophoresis and electrophoretic mobility shift assay (data not shown).

One of the most interesting observations in this study results from the identification of numerous nuclear proteins and also qualitative changes of protein pattern evaluated in two stages during early development and between two closely related species.

This work provides the identification of stage- or species-specific proteins as well as common ones, independently off protein abundance variations. This exploratory approach reveals a number of proteins potentially involved in the regulation/control of protein synthesis in the grain and also several proteins involved in development. Above all, the study enabled the transition from qualitative to quantitative nuclear proteomics. Hence, a major task will now be the use of quantitative proteomic approaches to correlate variations in the relative abundances and post-translational states of nuclear proteins in wheat grains across developmental stages including or not environmental and genetic fluctuations. The study of *T. monococcum*, because of its diploid genome, should greatly fasten our understanding of regulations mechanisms in bread wheat.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00446.

Figure S1: Graphical representation of the functional categories of proteins found in wheat grain nuclei enriched fraction. Figure S2: Classification in functional categories obtained for developing grains of *T. aestivum* and *T. monococcum* at 150 and 250 °Cd after anthesis. (PDF)

Table S1: Proteins identified in *T. aestivum* and *T. monococcum* grain nuclei at 150 and 250 °Cd after anthesis. (XLSX)

AUTHOR INFORMATION

Corresponding Author

*Tel: + 33 (0)473 624 3529. E-mail: emmanuelle.bancel@clermont.inra.fr.

Present Address

⊥P.M.: INRA, UMR0759 Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux, 2 place Viala, F-34 060 Montpellier, France.

Author Contributions

‖E.B. and T.B. contributed equally to this manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Christelle Damon (Blaise Pascal University, Clermont-Ferrand) for her help with microscopic observation and Dr. Karine Gallardo (INRA, Dijon) for providing some of the antibodies used in this work. This work was supported in the context of a Ph.D. grant from the French Ministry for Higher Education and Research to T.B. and funding from the French Government managed by the Research National Agency (ANR) in the framework of the Investments for the Future (ANR-10-BTBR-03).

REFERENCES

(1) 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*, 1594−1611.

(2) Nadaud, I.; Girousse, C.; Debiton, C.; Chambon, C.; Bouzidi, M. F.; Martre, P.; Branlard, G. Proteomic and morphological analysis of early stages of wheat grain development. *Proteomics* **2010**, *10*, 2901−2910.

(3) Tasleem-Tahir, A.; Nadaud, I.; Chambon, C.; Branlard, G. Expression Profiling of Starchy Endosperm Metabolic Proteins at 21 Stages of Wheat Grain Development. *J. Proteome Res.* **2012**, *11*, 2754−2773.

(4) Bancel, E.; Rogniaux, H.; Debiton, C.; Chambon, C.; Branlard, G. Extraction and Proteome Analysis of Starch Granule-Associated Proteins in Mature Wheat Kernel (*Triticum aestivum* L.). *J. Proteome Res.* **2010**, *9*, 3299−3310.

(5) 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*, 371−379.
(6) Meziani, S.; Nadaud, I.; Gaillard-Martinie, B.; Chambon, C.; Benali, M.; Branlard, G. Proteomic analysis of the mature kernel aleurone layer in common and durum wheat. *J. Cereal Sci.* **2012**, *55*, 323−330.
(7) Meziani, S.; Nadaud, I.; Gaillard-Martinie, B.; Chambon, C.; Benali, M.; Branlard, G. Proteomic Comparison of the Aleurone Layer in Triticum Aestivum and Triticum Monococcum Wheat Varieties. *Curr. Proteomics* **2014**, *11*, 71−77.
(8) Narula, K.; Datta, A.; Chakraborty, N.; Chakraborty, S. Comparative analyses of nuclear proteome: extending its function. *Front. Plant Sci.* **2013**, *4*, 100.
(9) Bae, M. S.; Cho, E. J.; Choi, E. Y.; Park, O. K. Analysis of the *Arabidopsis* nuclear proteome and its response to cold stress. *Plant J.* **2003**, *36*, 652−663.
(10) Khan, M. M. K.; Komatsu, S. Rice proteomics: recent developments and analysis of nuclear proteins. *Phytochemistry* **2004**, *65*, 1671−1681.
(11) Tan, F.; Li, G. S.; Chitteti, B. R.; Peng, Z. H. Proteome and phosphoproteome analysis of chromatin associated proteins in rice (*Oryza sativa*). *Proteomics* **2007**, *7*, 4511−4527.
(12) Pandey, A.; Choudhary, M. K.; Bhushan, D.; Chattopadhyay, A.; Chakraborty, S.; Datta, A.; Chakraborty, N. The nuclear proteome of chickpea (*Cicer arietinum* L.) reveals predicted and unexpected proteins. *J. Proteome Res.* **2006**, *5*, 3301−3311.
(13) Repetto, O.; Rogniaux, H.; Firnhaber, C.; Zuber, H.; Kuster, H.; Larre, C.; Thompson, R.; Gallardo, K. Exploring the nuclear proteome of *Medicago truncatula* at the switch towards grain filling. *Plant J.* **2008**, *56*, 398−410.
(14) Li, G. S.; Nallamilli, B. R. R.; Tan, F.; Peng, Z. H. Removal of high-abundance proteins for nuclear subproteome studies in rice (*Oryza sativa*) endosperm. *Electrophoresis* **2008**, *29*, 604−617.
(15) Ferreira, B. S.; Branco, A. T.; de Oliveira, M. A.; Pereira, M. G.; de Souza Filho, G. Methodological improvements on extraction of nuclear proteins and its preliminary analysis during the maize (*Zea mays* L.) endosperm development. *Protein Pept. Lett.* **2006**, *13*, 981−984.
(16) Erhardt, M.; Adamska, I.; Franco, O. L. Plant nuclear proteomics - inside the cell maestro. *FEBS J.* **2010**, *277*, 3295−3307.
(17) Repetto, O.; Rogniaux, H.; Larre, C.; Thompson, R.; Gallardo, K. The grain nuclear proteome. *Front. Plant Sci.* **2012**, *3*, 289.
(18) Brocklehurst, P. A. Factors controlling grain weight in wheat. *Nature* **1977**, *266*, 348−349.
(19) Ugarte, C.; Calderini, D. F.; Slafer, G. A. Grain weight and grain number responsiveness to pre-anthesis temperature in wheat, barley and triticale. *Field Crops Res.* **2007**, *100*, 240−248.
(20) Nicolas, M. E.; Gleadow, R. M.; Dalling, M. J. Effect of post-anthesis drought on cell division and starch accumulation in developing wheat grains. *Ann. Bot* **1985**, *55*, 433−444.
(21) Aki, T.; Yanagisawa, S. Application of Rice Nuclear Proteome Analysis to the Identification of Evolutionarily Conserved and Glucose-Responsive Nuclear Proteins. *J. Proteome Res.* **2009**, *8*, 3912−3924.
(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*, 248−254.
(23) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25−29.
(24) Blum, T.; Briesemeister, S.; Kohlbacher, O. MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinf.* **2009**, *10*, 274.
(25) Brameier, M.; Krings, A.; MacCallum, R. M. NucPred - Predicting nuclear localization of proteins. *Bioinformatics* **2007**, *23*, 1159−1160.
(26) 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.
(27) Briesemeister, S.; Rahnenfuhrer, J.; Kohlbacher, O. YLoc-an interpretable web server for predicting subcellular localization. *Nucleic Acids Res.* **2010**, *38*, W497−W502.
(28) Goldberg, T.; Hecht, M.; Hamp, T.; Karl, T.; Yachdav, G.; Ahmed, N.; Altermann, U.; Angerer, P.; Ansorge, S.; Balasz, K.; Bernhofer, M.; Betz, A.; Cizmadija, L.; Do, K. T.; Gerke, J.; Greil, R.; Joerdens, V.; Hastreiter, M.; Hembach, K.; Herzog, M.; Kalemanov, M.; Kluge, M.; Meier, A.; Nasir, H.; Neumaier, U.; Prade, V.; Reeb, J.; Sorokoumov, A.; Troshani, I.; Vorberg, S.; Waldraff, S.; Zierer, J.; Nielsen, H.; Rost, B. LocTree3 prediction of localization. *Nucleic Acids Res.* **2014**, *42*, W350−W355.
(29) 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*, 2172−2185.
(30) Castillo, J.; Rodrigo, M. I.; Marquez, J. A.; Zuniga, A.; Franco, L. A pea nuclear protein that is induced by dehydration belongs to the vicilin superfamily. *Eur. J. Biochem.* **2000**, *267*, 2156−2165.
(31) Zhang, S.; Xia, X. F.; Shen, J. C.; Zhou, Y.; Sun, Z. R. DBMLoc: a database of proteins with multiple subcellular localizations. *BMC Bioinf.* **2008**, *9*, 127.
(32) Cruz, J. R.; Moreno Diaz de la Espina, S. M. D. Subnuclear compartmentalization and function of actin and nuclear Myosin I in plants. *Chromosoma* **2009**, *118*, 193−207.
(33) Wilson, K. L.; Dawson, S. C. Functional evolution of nuclear structure. *J. Cell Biol.* **2011**, *195*, 171−181.
(34) Denison, F. C.; Paul, A. L.; Zupanska, A. K.; Ferl, R. J. 14−3-3 proteins in plant physiology. *Semin. Cell Dev. Biol.* **2011**, *22*, 720−727.
(35) Kumar, R.; Kumar, A.; Subba, P.; Gayali, S.; Barua, P.; Chakraborty, S.; Chakraborty, N. Nuclear phosphoproteome of developing chickpea grainlings (*Cicer arietinum* L.) and protein-kinase interaction network. *J. Proteomics* **2014**, *105*, 58−73.
(36) Dahan, J.; Wendehenne, D.; Ranjeva, R.; Pugin, A.; Bourque, S. Nuclear protein kinases: still enigmatic components in plant cell signalling. *New Phytol.* **2010**, *185*, 355−368.
(37) Heazlewood, J. L.; Whelan, J.; Millar, A. H. The products of the mitochondrial orf25 and orfB genes are Fo components in the plant F1Fo ATP synthase. *FEBS Lett.* **2003**, *540*, 201−205.
(38) Lee, S. M.; Kim, J. H.; Cho, E. J.; Youn, H. D. A nucleocytoplasmic malate dehydrogenase regulates p53 transcriptional activity in response to metabolic stress. *Cell Death Differ.* **2009**, *16*, 738−748.
(39) Anderson, L. E.; Wang, X. W.; Gibbons, J. T. 3 enzymes of carbon metabolism or their antigenic analogs in pea leaf nuclei. *Plant Physiol.* **1995**, *108*, 659−667.
(40) Zheng, L.; Roeder, R. G.; Luo, Y. S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell* **2003**, *114*, 255−266.
(41) Kim, J. W.; Dang, C. V. Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* **2005**, *30*, 142−150.
(42) Agrawal, G. K.; Bourguignon, J.; Rolland, N.; Ephritikhine, G.; Ferro, M.; Jaquinod, M.; Alexiou, K. G.; Chardot, T.; Chakraborty, N.; Jolivet, P.; Doonan, J. H.; Rakwal, R. Plant organelle proteomics: collaborating for optimal cell function. *Mass Spectrom. Rev.* **2011**, *30*, 772−853.

DOI: 10.1021/acs.jproteome.5b00446
J. Proteome Res. 2015, 14, 4432−4439

Supplementary information

Table S1: Proteins identified in *T. aestivum* and *T. monococcum* grain nuclei at 150 and 250°Cd after anthesis. Accession corresponding to Uniprot database. The proteins were sorted according to functional categories. Five predicting tools were used to determine nuclear localization (N), the number 1 to 5 correspond to the number of tools giving a nuclear localization.
(You can see this table online at: http://pubs.acs.org/doi/suppl/10.1021/acs.jproteome.5b00446)

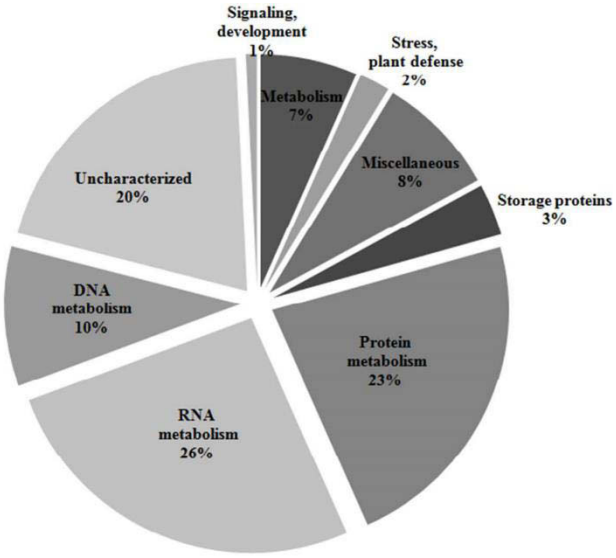


Figure S1: Graphical representation of the functional categories of proteins found in wheat grain nuclei enriched fraction. The 528 unique identified proteins were classified in 9 functional classes according to Gene ontology. The percentage of proteins in each functional class is shown.

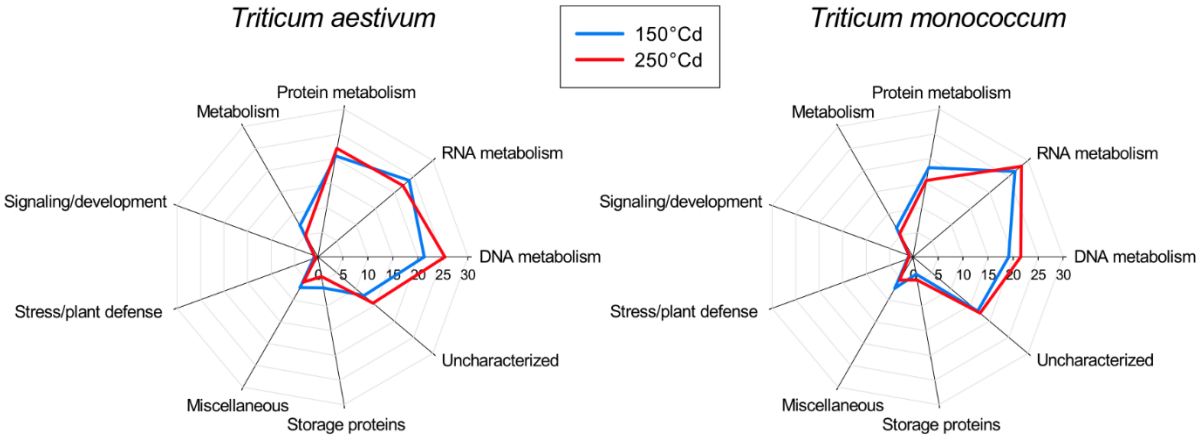


Figure S2. Classification in functional categories obtained for developing grains of *T. aestivum* and *T. monococcum* at 150 and 250C°Cdays after anthesis. The radar plots show the percentage of proteins in each functional class.

3. CONCLUSIONS

Dans cet article, les analyses réalisées ont permis de répondre à deux objectifs principaux : la validation de la méthode d'extraction des protéines nucléaires, et pour la première fois l'exploration du protéome nucléaire du grain de blé (Figure 21). Les colorations de noyaux ont révélé leur intégrité à l'issue de l'étape de purification sur gradient de densité. Les analyses par western blot ont ensuite mis en évidence l'enrichissement des extraits nucléaires en protéines histones, et l'absence dans ces extraits des protéines marqueurs des autres compartiments subcellulaires (vacuole, chloroplaste, mitochondrie, cytoplasme). A l'issue de l'analyse des protéines nucléaires en spectrométrie de masse, 797 protéines ont été identifiées. Parmi elles, 368 (46%) étaient communes aux différents échantillons, c'est-à-dire aux deux espèces de blé (*T. aestivum* et *T. monococcum*) et aux deux stades de développement étudiés (150 et 250°Cj après floraison). Sur cet aspect, nous avons pu voir que malgré les différences observées entre échantillons, les proportions des différentes fonctions retrouvées étaient comparables, notamment entre les deux stades de développement du grain d'une même espèce. Nous avons estimé grâce à des outils de localisation subcellulaire que parmi l'ensemble des protéines identifiées, 64 à 85% sont prédites comme ayant une localisation nucléaire. La méthode utilisée est donc robuste et permet d'obtenir des extraits protéiques enrichis en protéines nucléaires, pour les deux espèces de blé considérées.

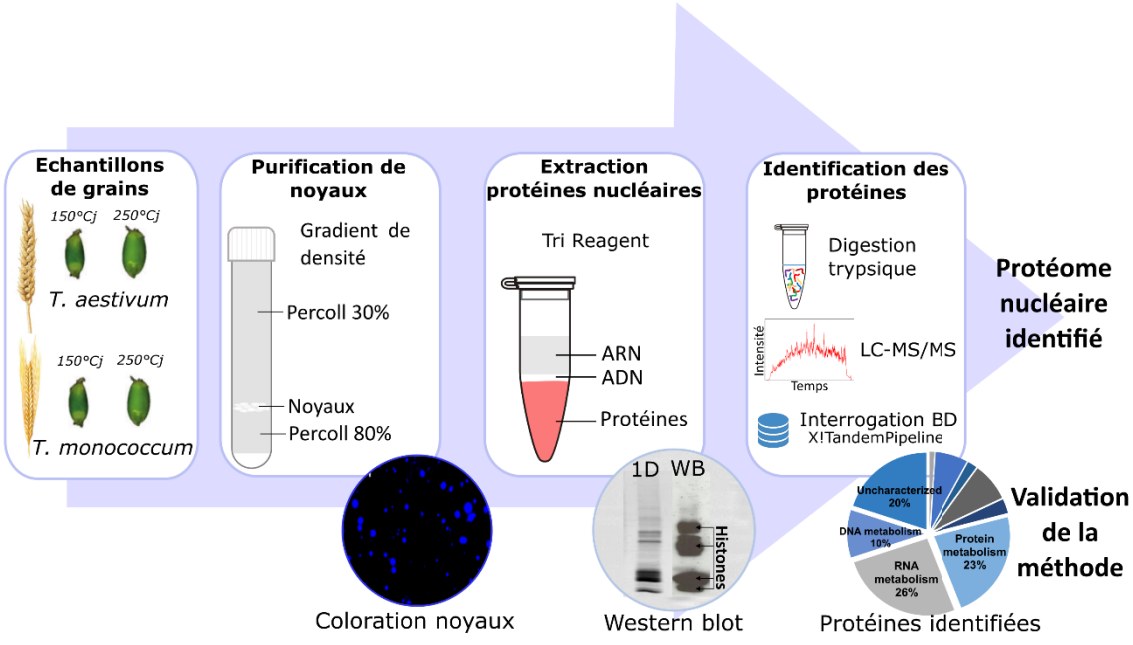


Figure 21. Principales étapes de la validation de la méthode d'extraction des protéines nucléaires.

En plus de valider la méthode, l'identification des protéines a permis de révéler l'identité d'un grand nombre d'acteurs nucléaires du grain de blé. Ces protéines sont impliquées principalement dans le métabolisme des ARN, des protéines et de l'ADN avec des fonctions associées aux processus d'assemblage des ribosomes, de la transcription, la maturation des ARNm ou encore la conformation de l'ADN.

Les différences observées entre les temps thermiques après floraison (150 et 250°Cj) d'un point de vue qualitatif ont laissé entrevoir de probables variations quantitatives au cours du développement du grain au niveau de ce sous-protéome. De plus, un rendement d'extraction de l'ordre de 110 à 160 µg de protéines extraites par gramme de grains frais, suivant l'espèce de blé et le stade de développement testés ainsi que la bonne reproductibilité de la méthode nous laissent la possibilité d'étudier ces variations quantitatives.

CHAPITRE 3 :
Le protéome nucléaire du grain en développement de blé tendre

1. INTRODUCTION

La compréhension des mécanismes qui contrôlent le développement du grain de blé est importante pour améliorer les paramètres de rendement et maintenir la qualité du grain. Avant d'atteindre la maturité physiologique, le grain passe par trois grandes phases : cellularisation, remplissage où les réserves sont accumulées et maturation, où le grain se déshydrate (Shewry *et al.*, 2012). Au niveau moléculaire, le développement du grain est caractérisé par la variation d'expression de nombreux gènes (Wan *et al.*, 2008; Capron *et al.*, 2012; Yu *et al.*, 2016).

Dans cette étude, nous avons émis l'hypothèse que les changements morphologiques et physiologiques du grain de blé observés au cours de son développement impliquent des variations d'abondance de protéines nucléaires à des moments spécifiques. L'objectif a ainsi été d'analyser le protéome nucléaire à différents temps thermiques après floraison pour visualiser ces variations quantitatives et identifier les protéines concernées.

Pour cela, nous avons choisi d'analyser le protéome nucléaire du grain de blé tendre à six temps thermiques après floraison, qui couvrent les trois phases du développement du grain et leurs transitions (Figure 22). Des grains ont ainsi été prélevés pendant la phase de cellularisation (150°Cj), à la transition entre les phases de cellularisation et de remplissage (250°Cj), pendant le remplissage (350 et 450°Cj), à la transition entre les phases de remplissage et de maturation (600°Cj) et pendant la maturation (750°Cj). Les protéines nucléaires ont été extraites de ces grains avec la méthode décrite dans le chapitre 2. Nous avons alors fait le choix de séparer ces protéines sur gel bidimensionnel. Afin de maximiser la résolution de la séparation, deux gradients de point isoélectrique (4-7 et 6-11) ont été analysés. Afin de révéler l'identité des protéines nucléaires variant ou non au cours du développement du grain, un maximum de spots protéiques a été analysé sans *a priori* par spectrométrie de masse.

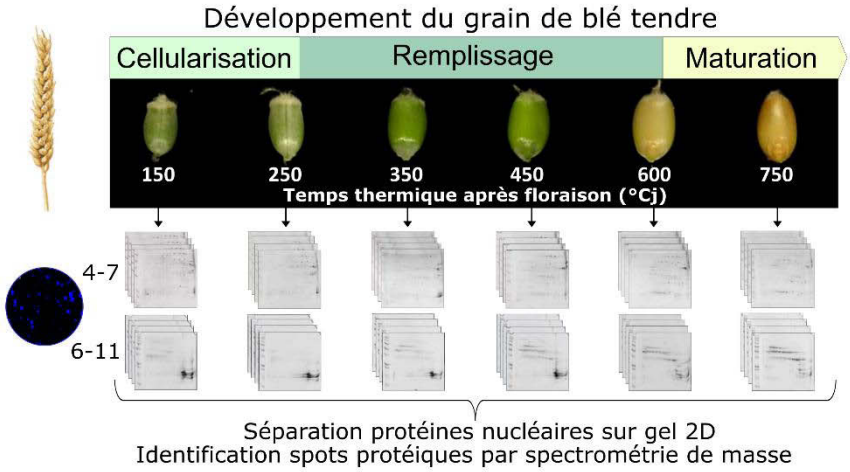


Figure 22. Analyse du protéome nucléaire du grain de blé tendre en développement. Les protéines nucléaires extraites des grains à six temps thermiques après floraison ont été analysées sur gel bidimensionnel (2D) dans deux gradients de point isoélectrique (4-7 et 6-11).

Références

Capron, D., Mouzeyar, S., Boulaflous, A., Girousse, C., Rustenholz, C., Laugier, C., Paux, E. and Bouzidi, M. (2012) Transcriptional profile analysis of E3 ligase and hormone-related genes expressed during wheat grain development. *BMC Plant Biol.*, **12**, 35.

Shewry, P.R., Mitchell, R. a. C., Tosi, P., et al. (2012) An integrated study of grain

development of wheat (cv. Hereward). *J. Cereal Sci.*, **56**, 21–30.

Wan, Y., Poole, R.L., Huttly, A.K., et al. (2008) Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics*, **9**, 121.

Yu, Y., Zhu, D., Ma, C., Cao, H., Wang, Y., Xu, Y., Zhang, W. and Yan, Y. (2016) Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *Crop J.*, **4**, 92–106.

2. ARTICLE 2 : Changes in the nuclear proteome of developing wheat (*Triticum aestivum* L.) grain

Titouan Bonnot[1, 2], Emmanuelle Bancel[1, 2], Christophe Chambon[3], Julie Boudet[1, 2], Gérard Branlard[1, 2] and Pierre Martre[1, 2]

[1]INRA, UMR1095 Génétique, Diversité et Ecophysiologie des Céréales, 5 chemin de Beaulieu, F-63 039 Clermont-Ferrand, France

[2]Université Blaise Pascal, UMR1095 Génétique, Diversité et Ecophysiologie des Céréales, Avenue des Landais, F-63 170 Aubière, France

[3]Plateforme d'Exploration du Métabolisme, INRA, Saint-Genès Champanelle, France

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 October 2015
doi: 10.3389/fpls.2015.00905

Changes in the nuclear proteome of developing wheat (*Triticum aestivum* L.) grain

Titouan Bonnot[1,2], *Emmanuelle Bancel*[1,2]*, *Christophe Chambon*[3], *Julie Boudet*[1,2], *Gérard Branlard*[1,2] and *Pierre Martre*[1,2†]

[1] UMR1095 Genetics, Diversity and Ecophysiology of Cereals, Institut National de la Recherche Agronomique, Clermont-Ferrand, France, [2] UMR1095 Genetics, Diversity and Ecophysiology of Cereals, Blaise Pascal University, Aubière, France, [3] Metabolism Exploration Platform Proteomic Component, Institut National de la Recherche Agronomique, Saint-Genès Champanelle, France

OPEN ACCESS

Edited by:
Dominique Job,
Centre National de la Recherche Scientifique, France

Reviewed by:
Ombretta Repetto,
Centro di Riferimento Oncologico di Aviano—SOC FSC, Italy
Beata Petrovska,
Centre of Plant Structural and Functional Genomics, Czech Republic

***Correspondence:**
Emmanuelle Bancel
emmanuelle.bancel@clermont.inra.fr

†Present Address:
Pierre Martre,
UMR759 Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux, Institut National de la Recherche Agronomique, Montpellier, France

Specialty section:
This article was submitted to Plant Proteomics, a section of the journal Frontiers in Plant Science

Received: 01 August 2015
Accepted: 10 October 2015
Published: 28 October 2015

Citation:
Bonnot T, Bancel E, Chambon C, Boudet J, Branlard G and Martre P (2015) Changes in the nuclear proteome of developing wheat (Triticum aestivum L.) grain. Front. Plant Sci. 6:905. doi: 10.3389/fpls.2015.00905

Wheat grain end-use value is determined by complex molecular interactions that occur during grain development, including those in the cell nucleus. However, our knowledge of how the nuclear proteome changes during grain development is limited. Here, we analyzed nuclear proteins of developing wheat grains collected during the cellularization, effective grain-filling, and maturation phases of development, respectively. Nuclear proteins were extracted and separated by two-dimensional gel electrophoresis. Image analysis revealed 371 and 299 reproducible spots in gels with first dimension separation along pH 4–7 and pH 6–11 isoelectric gradients, respectively. The relative abundance of 464 (67%) protein spots changed during grain development. Abundance profiles of these proteins clustered in six groups associated with the major phases and phase transitions of grain development. Using nano liquid chromatography-tandem mass spectrometry to analyse 387 variant and non-variant protein spots, 114 different proteins were identified that were classified into 16 functional classes. We noted that some proteins involved in the regulation of transcription, like HMG1/2-like protein and histone deacetylase HDAC2, were most abundant before the phase transition from cellularization to grain-filling, suggesting that major transcriptional changes occur during this key developmental phase. The maturation period was characterized by high relative abundance of proteins involved in ribosome biogenesis. Data are available via ProteomeXchange with identifier PXD002999.

Keywords: wheat, developing grain, nuclear proteins, 2D gel electrophoresis, LC-MS/MS

INTRODUCTION

Wheat (*Triticum aestivum* L.) grain is a major staple crop in many parts of the world. The end-use value is determined by complex molecular interactions that occur during grain development. Development of wheat grain is typical of grass seeds and is commonly subdivided into three developmental phases that overlap (Sabelli and Larkins, 2009). After double fertilization, the triploid endosperm divides successively without cytokinesis leading at 70°Cd after anthesis (i.e., 3–4 days after anthesis at an average daily temperature of 20°C) to the formation of a coenocyte whose nuclei are distributed throughout the endosperm (Mares et al., 1975). Cellularization follows which is a phase of cell division and differentiation until 220°Cd (11 days) after anthesis (Chojecki et al., 1986). The effective grain-filling phase follows when storage compounds, mainly starch

and proteins, rapidly accumulate (Shewry et al., 2012). The rate of accumulation of starch and proteins slows down at around 550°Cd (27 days) after anthesis, when endosperm nuclei and protein bodies become compressed by starch granules (Hoshikawa, 1962; Ferreira et al., 2012) and progressively disintegrate. Accumulation stops at 650–700°Cd (32–35 days) after anthesis when the concentration of water in grain is close to 45 g per 100 g of fresh mass (Schnyder and Baum, 1992). Grains then start a phase of rapid desiccation and maturation during which desiccation tolerance is acquired. Because of the importance of wheat grain in the human diet, much research has focused on identifying processes which regulate these different phases of development in order to optimize grain yield and its quality (Shewry et al., 2012).

The regulation of most of these processes involves transcriptional regulation and the nucleus plays a key role in the regulation of grain development and storage compound accumulation. In plants, the nuclear proteome of leaves or whole seedlings has been studied for several species (Erhardt et al., 2010; Petrovská et al., 2015) including cereals like *Oryza sativa* (Khan and Komatsu, 2004; Tan et al., 2007; Aki and Yanagisawa, 2009; Choudhary et al., 2009; Jaiswal et al., 2013), *Hordeum vulgare* (Petrovská et al., 2014), and *Zea mays* (Ferreira et al., 2006; Guo et al., 2014). However, there have been few such studies on seeds (Repetto et al., 2012). In *O. sativa*, 468 nuclear proteins were identified from endosperm at 9 days after pollination (dap) (Li et al., 2008) and in *Medicago truncatula* 143 different nuclear proteins were identified from whole seeds harvested at 12 dap (Repetto et al., 2008). A study of *Z. mays* showed that some nuclear proteins extracted from endosperm isolated from grains harvested between 8 and 35 dap, analyzed on one-dimensional (1D) gels, were more abundant at certain times of development (Ferreira et al., 2006), but these proteins remain to be identified. No proteomic study has analyzed the temporal changes in abundance of nuclear proteins during grain development. However, identifying and quantifying nuclear proteins is an important step in characterizing some of the numerous regulatory mechanisms that take place during the dynamic phases of grain development. We hypothesized that the developmental physiology and morphology of the wheat grain requires changes in abundance of several nuclear proteins at specific times of grain development.

The aim of the present study was to analyze the nuclear proteome of the developing wheat grain in order to obtain a first overview of which nuclear proteins vary in abundance during grain development. Nuclear proteins were extracted from wheat (*T. aestivum* L.) grains collected during the cellularization, effective grain-filling and maturation phases of development, and analyzed using two-dimensional (2D) gel electrophoresis and electrospray ionization ion trap mass spectrometry (ESI-IT-MS/MS). This allowed us to show that some nuclear proteins involved in signaling, proteolysis, transcription regulation or ribosome biogenesis were more abundant at specific developmental phases or phase transitions.

MATERIAL AND METHODS

Plant Material

Plants of hexaploid winter wheat (*T. aestivum* L.) cv Recital were used in this study. Seeds were sown in plug trays filled with a peat moss mixture and were kept in a greenhouse until the ligule of the third leaf appeared. Air temperatures in the greenhouse were maintained at 18/10°C (light/dark) and air relative humidity at 70/50% (light/dark). Plants were then vernalized for 8 weeks in a growth chamber where the air temperature was maintained at 4 ± 1°C, the air relative humidity at 40% and the mean daily photosynthetic photon flux density (PPFD) at the top of the plants at 43 mmol m^{-2} d^{-1} during the 8-h photoperiod. After vernalization, the plants were transplanted into 293-mL plastic pots filled with a mixture of soil-pozzolan (2:1, w/w) and transferred to a walk-in growth chamber. The conditions in the growth chamber were 20/15°C (light/dark), 55/75% air relative humidity (light/dark), with an average PPFD of 550 µmol m^{-2} s^{-1} at top of the plants during the 16-h photoperiod. Plants were irrigated twice a day with a commercial nutrient solution.

Air temperature at the top of the plants was measured continuously. Main stems were tagged when the anthers of the central florets emerged (anthesis date). The sum of mean daily air temperature after anthesis was calculated to follow grain development in thermal time in degree-days (°Cd) above 0°C after anthesis. Grains were harvested at 150, 250, 350, 450, 600, and 750°Cd after anthesis and stored at −80°C. Only grains of the first floret from the central part of the ears were collected (approximately 10 grains per ear). Four independent replicates were used.

Nuclei Isolation and Nuclear Protein Extraction from Wheat Grains

The method used in the present study to purify nuclei from wheat grains and to extract nuclear proteins was recently validated by Bancel et al. (2015). The authors verified the absence of contamination from non-nuclear proteins by western blots with primary antibodies which detect protein markers of the different sub-cellular compartments.

Briefly, nuclei were isolated from 2 g of grains. Grains were ground in extraction buffer [20 mM Hepes-KOH, pH 7, 5 mM $MgCl_2$, 10 mM 2-ME, 0.5 mM PMSF, 0.1% (v/v) phosphatase inhibitor cocktail (Sigma-Aldrich)] with a Polytron homogenizer (Kinematica POLYTRON® PT 10) during 1 min. The homogenate was filtered through two layers of Miracloth (Calbiochem) to remove cell debris and cells were lysed by adding 0.5% (v/v) Triton X-100. After incubation for 15 min at 4°C, the resulting lysate was centrifuged at 1000 × *g* for 10 min at 4°C. Each pellet was then washed four times by resuspension in 1 mL of extraction buffer followed by centrifugation at 1000 × *g* for 10 min at 4°C. Nuclei were purified from the pellet by centrifugation at 930 × *g* for 30 min at 4°C through a stepwise Percoll density gradient, 30–80% Percoll prepared in extraction buffer. Nuclei floating at the interface were collected and washed twice with 3 mL of extraction buffer followed by centrifugation at 3500 × *g* for 5 min at 4°C. To verify the purity of isolated nuclei, nuclei pellets were washed in 500 µL of PBS (137 mM NaCl,

2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4) and centrifuged at 3500 × *g* for 5 min at 4°C. The supernatant was removed and nuclei were stained in phosphate buffered saline (PBS) solution containing 0.1 μg/mL Hoechst for 5 min in the dark. After two washes in 50 μL of PBS, 5 μL aliquots were observed under fluorescence microscopy (Zeiss Axioplan 2 microscope). To verify absence of pigments in nuclei pellets, a chlorophyll assay was performed according to (Pandey et al., 2006). Briefly, 100 μL of sample was mixed with 100 μL of water and 800 μL of acetone. After centrifugation at 1000 × *g* for 5 min, the optical density was measured at 652 nm. The amount of chlorophyll was observed as μg per μL by calculating optical density/34.5 (at 652 nm chlorophyll a and b intersect, 34.5 is the specific absorption coefficient for both pigments at this wavelength).

Nuclear proteins were prepared using TRI Reagent® (Sigma-Aldrich) according to the manufacturer's instructions. The final protein pellet was dried under ambient conditions for 1 h and then stored at −20°C.

SDS-PAGE Analysis and Immunoblotting Using Anti-histone H3 Antibody

Nuclear protein pellets and proteins from supernatants (S1–S5) collected during nuclei purification were solubilized in 100 μL of solubilization buffer (45 mM Tris-HCl, pH 6.8, 50 mM DTT, 1% (w/v) SDS, 10% (v/v) glycerol, 0.001% (w/v) bromophenol blue). A fixed volume (35 μL of each supernatant or 25 μL of final nuclear protein extract) was loaded onto 12.5% SDS-polyacrylamide gels as described in (Bancel et al., 2015). For immunoblotting analysis, proteins were transferred from the 1D gels to nitrocellulose membranes (Hybond ECL, GE Healthcare) during 1 h in a semidry unit apparatus (GE Healthcare). Membranes were incubated with an anti-histone H3 antibody (Abcam) diluted at 1:1000. Membranes were then incubated with anti-rabbit secondary antibody coupled to horseradish peroxidase (HRP, GE Healthcare) diluted at 1:5000. The chemiluminescence was developed according to the manufacturer's instructions (ECL Western Blotting, SuperSignal West Pico Chemiluminescent Substrate kit, Amersham).

Two-dimensional Electrophoresis of Nuclear Proteins

Dried pellets containing nuclear proteins were dissolved in 50 μL of 2D gel sample buffer [7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 70 mM dithiothreitol, 1% (v/v) immobilized pH gradient (IPG) buffer (either for the pH 4–7 range or the pH 6–11 range), and 0.34% (v/v) protease inhibitor (Sigma-Aldrich)] for 1 h at room temperature with constant agitation. An aliquot (5 μL) of each sample was used to quantify protein content (Bradford, 1976) using bovine serum albumin as standard. Isoelectric focusing was carried out with 150 μg of proteins, made up to 250 μL with the 2D buffer containing 0.05% (w/v) bromophenol blue and used to passively rehydrate 13-cm immobilized pH gradient strips (pH range 4–7 or 6–11 Immobilin Dry Strips, GE Healthcare) overnight at 20°C. For the comigration gels, made to be a reference for the digital analysis step, isoelectric focusing was performed with 240 μg of proteins (40 μg of nuclear proteins from each of the six thermal times after anthesis). Isoelectric focusing was performed for a total of 60,000 voltage hours (Vhr) on a IPGphor II apparatus (GE Healthcare). Focused proteins on strips were then reduced with 2% (w/v) dithiothreitol in 0.1 M Tris-HCl buffer (pH 8.8) containing 6 M urea, 30% (v/v) glycerol, and 2% (w/v) SDS for 15 min, followed by alkylation with 2.5% (w/v) iodoacetamide in the same buffer for 15 min. The strips were then loaded onto 12.5% polyacrylamide gels for SDS-PAGE separation in the second dimension. The migration conditions were 10 mA per gel for the first 30 min, then 35 mA per gel for 2.5 h. Gels were stained using Coomassie Brilliant Blue G250 (CBB, Sigma-Aldrich) (Neuhoff et al., 1985). To improve detection of low abundance protein spots and allow their collection, gels were destained overnight in a solution containing 40% (v/v) ethanol and 10% (v/v) acetic acid and silver-stained following a mass spectrometry compatible method (Shevchenko et al., 1996).

Image and Statistical Analyses

Images (300 dpi, 16-bit greyscale pixel depth) of two-dimensional gels stained with CBB were acquired with a GS-800 (Biorad) scanner and analyzed using SameSpots v4.5 (TotalLab) 2D gel image analysis software. Statistical analyses were performed on normalized protein spot volume values. Differences in normalized protein spot volume due to grain development were analyzed using One-way ANOVA. *P*-values and adjusted *P*-values (*q*-value) were calculated using SameSpots procedures. The abundance of a protein spot was considered to have changed during grain development when its *P*-value and *q*-value were both <0.05. In this case, the protein spot was considered as "variant" and others as "non variant." Principal component analysis was performed using the FactoMineR (Husson et al., 2014) package for R v3.0.1 (R Core Team, 2013) statistical software on the set of spots detected on 2D gels and hierarchical clustering on principal components was computed on significant spots to build protein abundance profiles.

Protein Identification

Protein spots were excised manually from 2D gels. For 2D gels stained with CBB, protein spots were destained once with 25 mM NH_4HCO_3 containing 5% (v/v) acetonitrile (ACN) for 30 min and twice with 25 mM NH_4HCO_3, 50% (v/v) ACN for 30 min. Spots were then dehydrated in 100% ACN for 10 min and dried for 15 min under an extraction hood at room temperature. For 2D gels stained with silver nitrate, protein spots were first destained with 30 mM $K_3Fe(CN)_6$, 100 mM $Na_2S_2O_3$ for 1–2 min, then washed twice in water for 15 min, before following the CBB destaining steps as above. Proteins were digested overnight at 37°C by adding 120 ng of trypsin (Promega). After extracting peptides with ACN, 8 μL of hydrolysate were injected into an Ultimate® 3000 HPLC system (Dionex) coupled to an electrospray ionization ion trap mass spectrometer (ESI-IT-MS/MS; LTQ Velos, ThermoScientific).

Protein identity was sought by using Mascot v2.3 (Matrix Science) software against a custom database containing 249,032 sequences from *Aegilops tauschii*, *A. thaliana*, *Brachypodium*

distachyon, *H. vulgare*, *O. sativa*, *T. aestivum*, and *T. urartu* and sequences from the wheat transcription factor database wDBFT (Romeuf et al., 2010). Proteins were considered to be identified if at least two non-redundant peptides were found to match a single reference in the databases. A cut-off was applied for individual peptide ion scores according to the significance threshold of the MASCOT program ($P < 0.05$). Curated protein sequences which had no functional information were submitted as BLASTP searches against the National Center for Biotechnology Information non-redundant database (http://blast.ncbi.nlm.nih.gov). Proteins were then classified in functional classes according to the KEGG PATHWAY database (Kanehisa et al., 2014) and gene ontology (Ashburner et al., 2000). Subcellular localization of identified proteins was predicted by *in silico* analysis using Multiloc2 (Blum et al., 2009), WolfPSort (Horton et al., 2007), Y loc (Briesemeister et al., 2010), and LocTree (Goldberg et al., 2014) programs. The top two hits were considered for Multiloc and Y loc. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaíno et al., 2014) via the PRIDE partner repository with the dataset identifier PXD002999.

RESULTS

Nuclei Purified and Nuclear Proteins Extracted from Developing Wheat Grains

Here we analyzed the nuclear proteome of developing wheat grain. For this, grain was harvested at six thermal times corresponding to the cellularization phase (150°Cd after anthesis), at the transition between cellularization and grain-filling phases (250°Cd after anthesis), during the grain-filling phase (350 and 450°Cd after anthesis), at the transition between grain-filling and maturation phases (600°Cd after anthesis), and during the maturation phase (750°Cd after anthesis). Nuclei were isolated from wheat grains using Percoll density gradient purification. The integrity of isolated nuclei was verified by Hoechst staining. For all stages of grain development studied, Hoechst staining showed that nuclei with an average diameter of approximately 20 μm had been purified (Supplementary Figure 1A). Chlorophyll assay was also performed which showed that nuclear extracts were free of these pigments (Supplementary Figure 1B). Nuclear proteins and proteins from supernatants collected during nuclei purification were analyzed by SDS-PAGE and western blot analysis with an anti-histone H3 antibody (**Figures 1A,B**). For each stage of grain development, 1D protein profiles of the nuclear fractions were different from those of supernatants (**Figure 1A**). After blotting, histone H3 proteins were clearly detected in the nuclear fraction, confirming that the protein extracts were enriched in nuclear proteins (**Figure 1B**). Histone H3 was less abundant in supernatant S1 from protein samples from 250, 350, 450, and 600°Cd after anthesis indicating that a few nuclei were lost during purification.

The 1D gels show that the quantity of nuclear proteins extracted from an equal mass of grain decreased during grain development. This result was confirmed by protein assays of the nuclear protein fraction (**Figure 1C**). However, the amount per grain of nuclear proteins extracted at various thermal times after anthesis were similar, except at 750°Cd after anthesis when the variability was too high to evaluate this (**Figure 1D**).

Two-dimensional Electrophoresis of Wheat Grain Nuclear Proteins

To maximize resolution of protein spots in 2D electrophoresis, two different pH gradients were used in the first dimension of isoelectric focusing. Comigration gels are shown in **Figure 2** in which samples from all 6 stages of wheat grain development were combined. With the pH 4–7 gradient, 371 spots were detected and with the pH 6–11 gradient 299 spots were detected. From this total of 690 protein spots, 213 were manually excised from the pH 4–7 gradient gel and 174 from the pH 6–11 gradient gel, without any a priori (**Figure 2**). From the total of 387 excised spots, 343 (88.6%) polypeptides were identified by LC-ESI-MS/MS, which correspond to 114 different proteins (Supplementary Tables S1, S2). In many cases, the same protein was found in multiple spots, indicative of isoforms and/or post-translational modifications.

The identified proteins were organized into 16 functional classes (**Figure 3** and Supplementary Table S1). Among the proteins with known functions, a large proportion (15%) corresponded to ribosomal proteins involved in ribosome biogenesis. The second largest category (11%) comprised proteins involved in transcription and transcription regulation. For example, DNA-directed RNA polymerases I, II, and III subunit rpabc3 is involved in RNA synthesis, RNA helicases, arginine/serine-rich splicing factor, and Mago nashi-like protein are all involved in mRNA processing. Transcriptional regulators were also identified such as a MADS-box transcription factor and a HMG1/2-like protein. Three spots corresponded to histone deacetylases, which mediate the deacetylation of lysine residues on the N-terminal part of core histones and three others matched the FACT complex subunit SSRP1-B which regulates transcription by modifying nucleosomal structure. In the same category, WD repeat-containing protein RBAP1 was also identified. Proteins involved in protein folding represented the third largest category (9%) among characterized proteins. Other proteins were related to translation (6%) including initiation and elongation factors. In the nucleosome assembly category (6%) histones H1, H2A, H2B, and H4 were identified. Finally, some proteins with storage functions were identified in the nutrient reservoir activity category (6%) and are thought to be contaminants from the purification method. The function of 17 (16%) proteins was not known so they were classified as uncharacterized.

The Relative Abundance of Many Nuclear Proteins Varied During Grain Development

For the six thermal times after anthesis, CBB stained gels from four biological replicates were analyzed by digital imaging. Principal component analysis was performed with the normalized volumes of the 690 protein spots detected by image analysis (Section Two-dimensional Electrophoresis of Wheat Grain Nuclear Proteins). The four replicates for a given

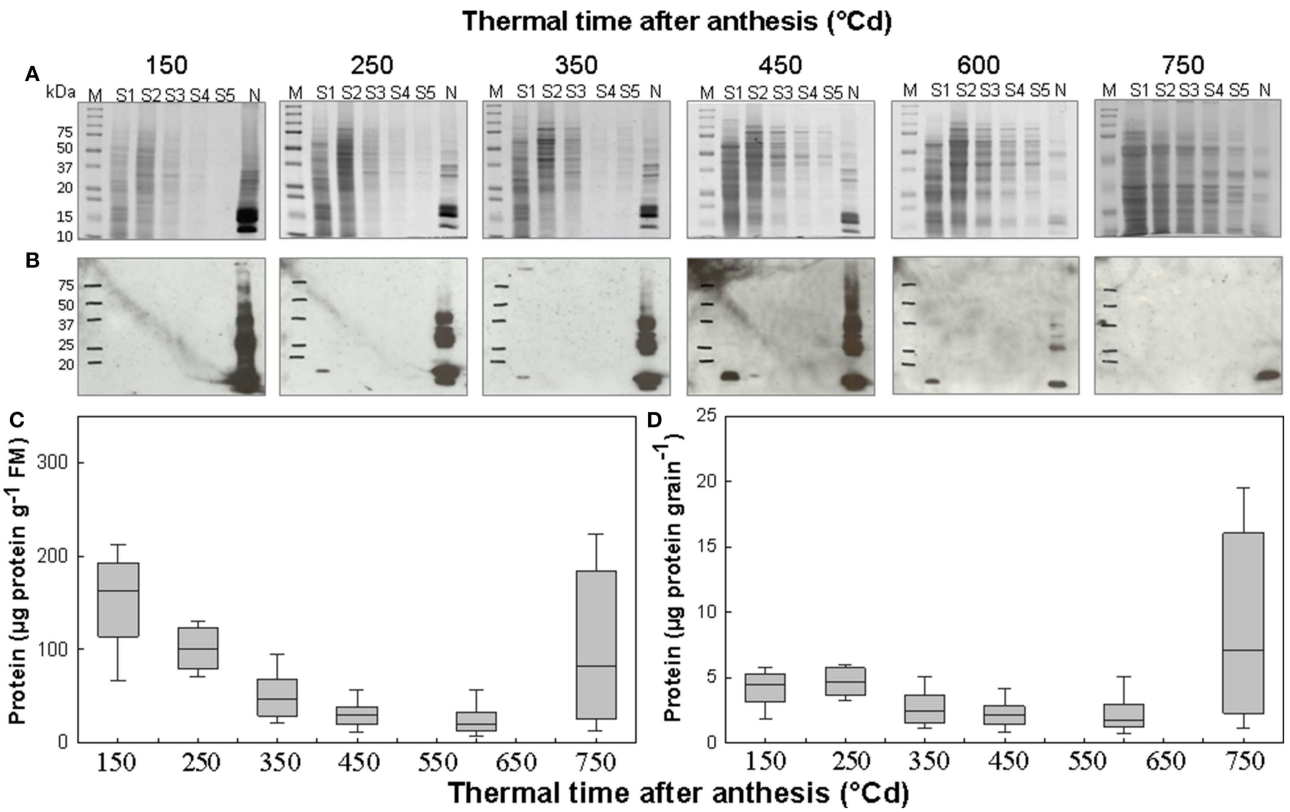


FIGURE 1 | Nuclear proteins extracted from developing wheat grains. Proteins from supernatants (S1–S5) and nuclear-enriched fractions (N) for each thermal time after anthesis (150–750°Cd) were analyzed on SDS-PAGE **(A)** then western blot analysis was performed using antibody directed against histone H3 **(B)**. M, molecular weight protein standards. Amount of proteins in the nuclear-enriched fraction per fresh mass (FM) of grain **(C)** or per grain **(D)** vs. thermal time after anthesis. In **(C,D)** boxes show the 25th to 75th percentile range, horizontal lines in boxes show medians, and error bars outside boxes show the 10th to 90th percentile range for $n \geq 10$ independent extractions.

time point segregated away from those of other time points (Supplementary Figure 2). From the 690 protein spots detected, 226 (33%) had a constant relative abundance during grain development, and 464 varied significantly (67%). Protein spots were excised without any a priori. In this way, among the 387 protein spots analyzed by LC-MS/MS, 153 (40%), corresponding to 69 different proteins, did not vary in relative abundance during grain development, and are qualified as non-variant, whereas 234 (60%), corresponding to 72 different proteins, were variant. An interesting initial conclusion is that a protein can be identified in multiple spots, some of which vary during grain development, and others which do not.

Proteins involved in functional classes ribosome biogenesis (12 proteins), uncharacterized (9 proteins) and transcription/transcription regulation (8 proteins) were the most numerous among the non-variant proteins (**Figure 4**). The ribosome biogenesis functional class (10 proteins) was also highly represented among the variant proteins (**Figure 4**). The relative abundance of seven histones varied significantly. Only 50% (6/12) of the identified proteins related to transcription regulation varied during grain development. Conversely, the functional classes of proteolysis and plant defense were more highly represented in the variant protein group (five different proteins from each class) than in the non-variant protein group (1 and 3 different proteins in each class, respectively).

The clustering analysis was first performed separately on the pH gradients. Since the two pH ranges gave similar clusters, the clustering analysis presented here was performed using data from the two pH ranges. The 464 variant protein spots detected by image analysis were grouped into six profiles according to their relative abundance at different stages of grain development (**Figure 5**). Profile 1 included 37 spots with a maximum normalized volume at 150°Cd after anthesis that decreased to a minimum value at 450°Cd after anthesis (**Figure 5A**). Profile 2 grouped 58 spots whose normalized volume peaked at 250°Cd after anthesis (**Figure 5B**). The 54 spots that defined profile 3 had a maximum normalized volume between 150 and 450°Cd after anthesis (**Figure 5C**). Profile 4 grouped 24 spots with normalized volume that peaked at 450°Cd after anthesis (**Figure 5D**). Profile 5 grouped 184 spots whose normalized volume increased throughout grain development (**Figure 5E**) and profile 6 included 107 spots whose normalized volume decreased from 150 to 450°Cd after anthesis and then increased until 750°Cd after anthesis (**Figure 5F**).

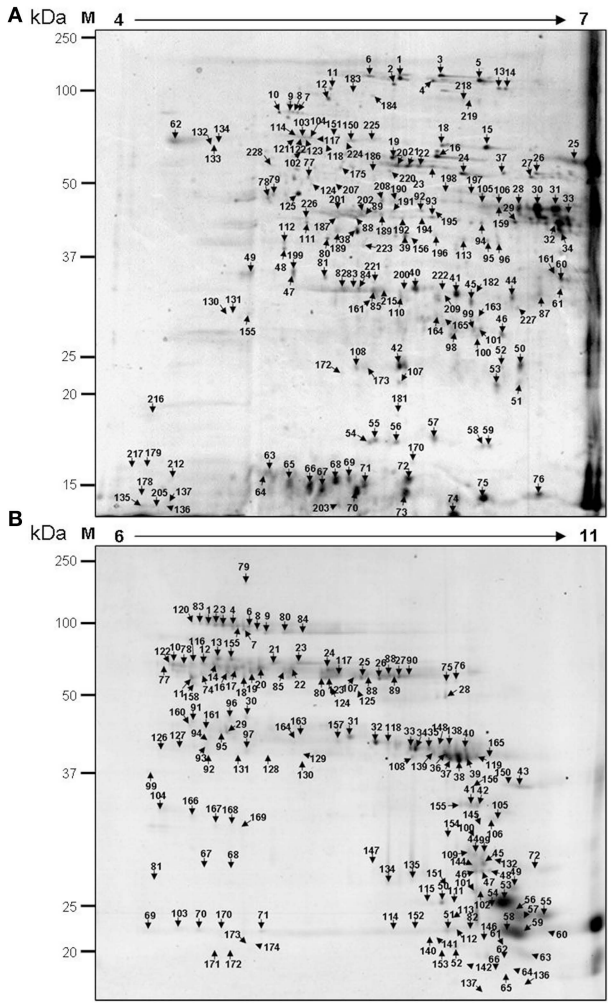


FIGURE 2 | Map of wheat grain nuclear protein spots. Comigration in 2-D electrophoresis of nuclear proteins extracted from grains harvested at six thermal time after anthesis. Proteins were loaded onto 13-cm IPG strips in the pH range 4–7 **(A)** and 6–11 **(B)** for isoelectric focusing in the first dimension and SDS-PAGE in the second dimension. M, molecular weight protein standards. Protein spots that were excised and analyzed by mass spectrometry are indicated.

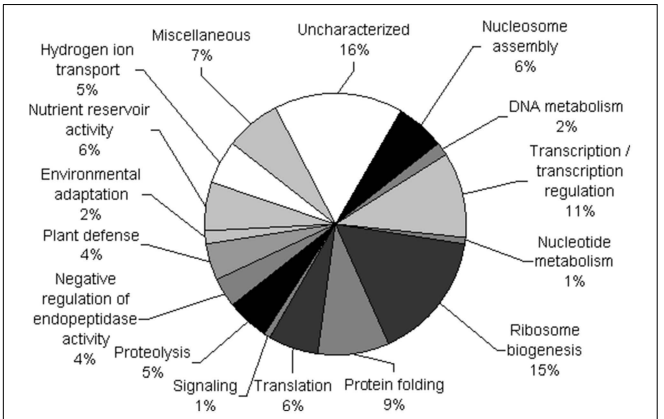


FIGURE 3 | Functional classification of nuclear proteins. The 114 different proteins identified were grouped into 16 functional classes according to the KEGG pathway database and gene ontology (GO). The percentage of proteins in each functional class is shown.

Each of the 72 identified variant proteins had at least one of the six profiles (Supplementary Table S1). There were 12 proteins with profile 1, 19 with profile 2, 15 with profile 3, 6 with profile 4, 21 with profile 5, and 21 with profile 6. In some cases therefore the same protein, perhaps different isoforms, had different abundance profiles. Profiles 1 and 2 were characterized by proteins involved in transcription and transcription regulation. For example, HMG1/2-like protein, Mago nashi-like protein and a histone deacetylase HDAC2 were more abundant during the cellularization phase of grain development (profile 1). Two helicase proteins and a HMG1/2-like protein peaked in relative abundance at the end of the cellularization phase (profile 2). Seven histones had profile 3 indicating that they were more abundant during the cellularization phase and the beginning of the grain-filling phase. Several variant ribosomal proteins had profiles 4, 5 and 6, and therefore were more abundant as a class during the latter part of the grain filling phase and during maturation. Proteins with profile 6, increasing steadily in abundance during the late grain-filling phase, included three serpins, which are known to negatively regulate endopeptidase activity, four proteins involved in protein folding and two proteins involved in plant defense.

Focus to Variation in Abundance for Some Identified Protein

In some cases, only one spot matched a unique protein and the volume of this spot varied significantly during seed development, e.g., the Mago nashi-like protein and the Do-like 9 protease with profile 1 (**Figures 6A,B**). In 25% of the cases (28 different proteins), several protein spots corresponding to the same protein did not vary significantly in the same way. For example, two protein spots corresponded to a histone deacetylase HDAC2. One spot varied as in profile 1, but the other did not vary during grain development (**Figure 6C**). This was also the case for a 60S acidic ribosomal protein P0 protein. One protein spot peaked in relative abundance at 600°Cd after anthesis, while the other spot remained constant (**Figure 6D**). In a few cases (16%), several protein spots corresponding to the same protein were variant but did not have the same relative abundance profile. For example, an HMG1/2-like protein was present in eight protein spots. Of these, two varied with profile 1 dynamics and four with profile 2 dynamics (**Figure 6E**). Nevertheless overall this protein was more abundant during the cellularization phase. Finally in some cases, a protein was identified in several spots which all showed the same abundance dynamics. For example, 40S ribosomal protein S12 was identified in three spots, which were all classified as profile 5 (**Figure 6F**), all increasing in abundance during grain development.

DISCUSSION

DNA replication, repair and modification, RNA transcription, and ribosome biogenesis occur in the nucleus. Continuous flux

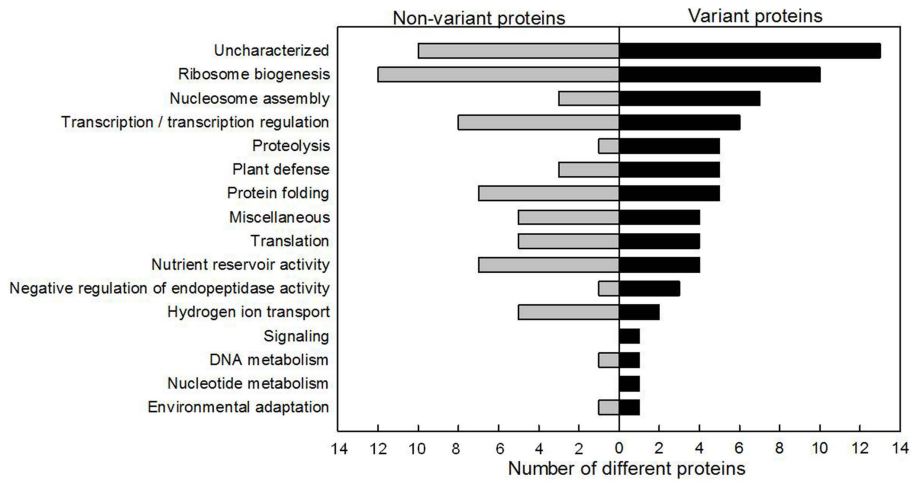


FIGURE 4 | Functional distribution of non-variant and variant nuclear proteins. Proteins identified showing no statistically significant change are represented on the left and the variant proteins are presented on the right.

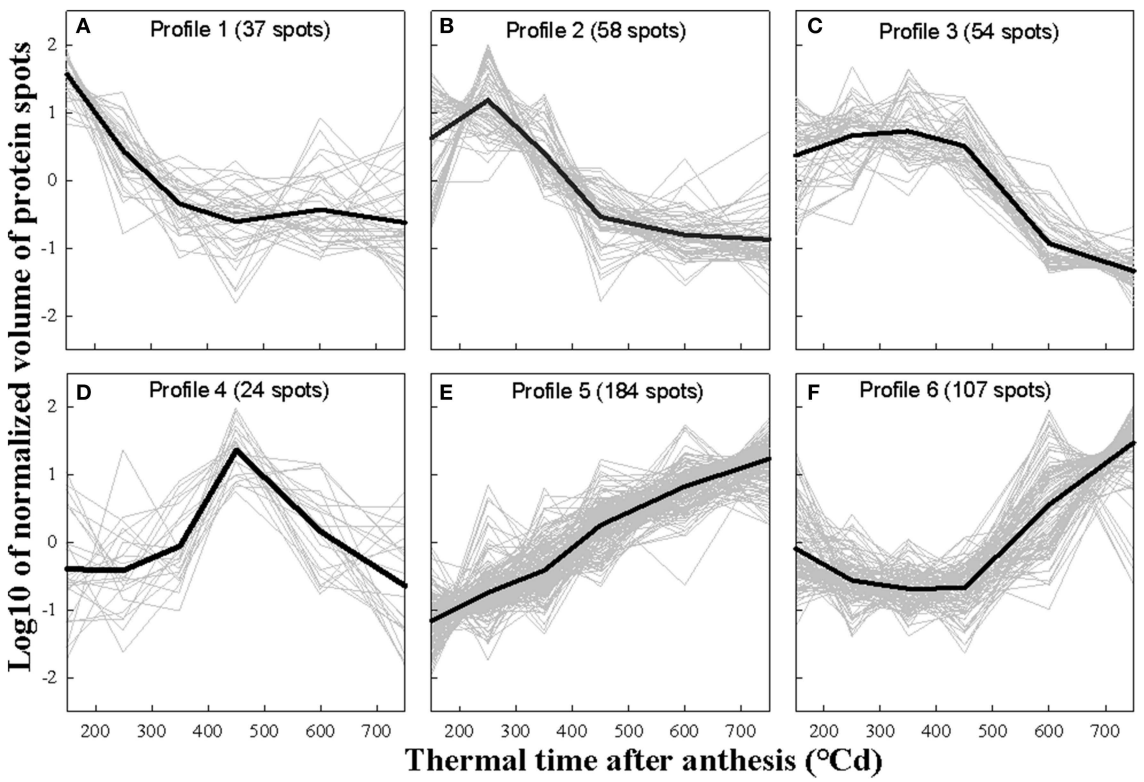


FIGURE 5 | Abundance profiles of variant protein spots. Six profiles **(A–F)** were established from the normalized volumes of the 464 variant spots using hierarchical clustering on principal components. Data are means of $n = 4$ replicates. In each profile, the black line represents the mean. The number of protein spots in each profile is given in parenthesis after the profile number.

of RNAs and proteins across the nuclear envelope make the nucleus a very dynamic structure and the site of developmental regulation. We analyzed the nuclear proteins that were extracted from nuclei at key stages of wheat grain development. Using 2D electrophoresis and LC-ESI-MS/MS, 114 different proteins were identified. Changes in the relative abundance of proteins and their isoforms during grain development are highlighted. Certain proteins which could have a potential role in the regulation of

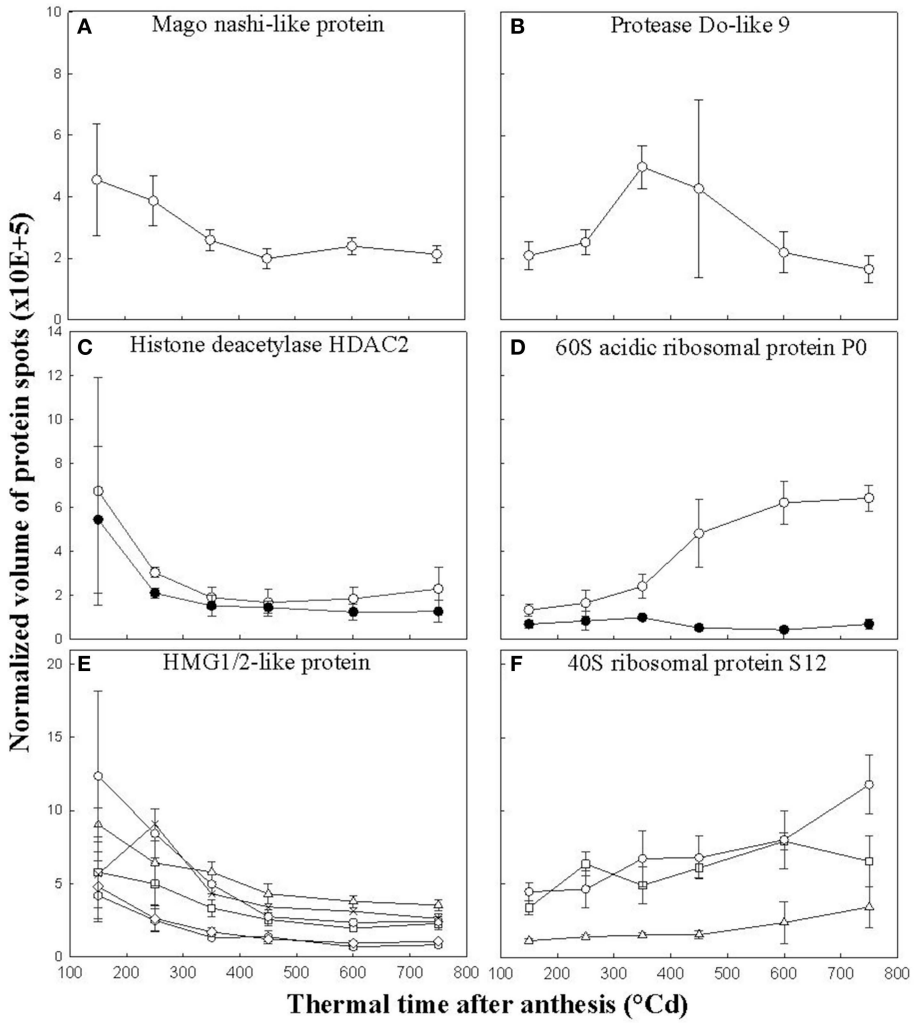


FIGURE 6 | Changes in volumes of several nuclear protein spots. Mago nashi-like protein **(A)** and protease Do-like 9 **(B)** were each identified in only one spot whose abundance varied significantly during grain development. Histone deacetylase HDAC2 **(C)** and 60S acidic ribosomal protein P0 **(D)** were each found in two spots including one which varied significantly (open circles) and one which did not vary significantly during grain development (closed circles). HMG1/2-like protein was identified in eight spots including six that varied, presented graphically in **(E)**. The 40S ribosomal protein S12 **(F)** was found in three spots whose volumes varied significantly during grain development. Data are means ± 1 s.d. for $n = 4$ independent replicates.

grain development are discussed, and could indeed be the object of further studies.

Nuclear Proteins Identified Provide Information on the Functions of Wheat Grain Nuclei

After translation in the cytoplasm, ribosomal proteins are transported to the nucleolus, where ribosome assembly begins (Fromont-Racine et al., 2003; Boisvert et al., 2007). Among the 114 identified proteins, 18 (15%) are ribosomal proteins. By comparison, in a study of the *M. truncatula* nuclear proteome, 32 ribosomal proteins (22%) were identified (Repetto et al., 2008). The second largest category of proteins (12 proteins) identified were those involved in transcription/transcription regulation. Some of them have been identified in previous studies of grain nuclear proteomes, such as the histone deacetylases (Li et al., 2008; Repetto et al., 2008). However, the HMG1/2-like protein, identified here in eight protein spots, had not previously been found in the nuclear proteome of grain from other species. The presence of histone proteins was expected as they are involved in nucleosome assembly, the first level of DNA compaction, and seven were indeed identified.

The presence of other proteins in the nucleus was harder to predict. Four different tools were used to predict subcellular localization. Seventy one percent of identified protein spots were predicted to correspond to a nuclear protein with at least two tools (N2, Supplementary Table S1), which suggest that these proteins correspond to nuclear actors or spend some time in the nucleus. However, proteins not predicted to be nuclear can't be excluded due to the limitation of the prediction tools. Indeed, several proteins known to spend some time in the nucleus were

not predicted to be nuclear with the tools employed here, like for example some ribosomal proteins, involved in the ribosome biogenesis that occur in the nucleolus. While it is probable that some of proteins identified are contaminants from the purification process, it has been estimated that 35% of total proteins have multiple subcellular locations (Zhang et al., 2008), so some are worth discussing. For example, luminal binding proteins and HSP70 act mainly in the endoplasmic reticulum or in the cytosol to contribute to the formation of three-dimensional structures of proteins or protein complexes. These two types of protein were found in wheat grain nuclei. HSP70 has also been identified in the nucleus of other plant species (Calikowski et al., 2003; Pandey et al., 2006; Repetto et al., 2008) and some HSP are translocated to the nuclei of hamster fibroblasts following heat stress (Nollen et al., 2001). Two guanine nucleotide binding proteins were identified. They are commonly associated with the plasmalemma acting in multiple signal transduction pathways. In eukaryotes, guanine nucleotide binding proteins can also be associated with endomembranes, nucleus and the cytoskeleton (Willard and Crouch, 2000). Four serpin proteins were identified in this study. In mammals, several members of the serpin family have been found to localize in the nucleus and some have a nuclear localization signal (Silverman et al., 2001).

Translation mostly takes place in the cytosol. However, initiation and elongation factors involved in translation have been identified in the nuclear proteomes of *M. truncatula*, *O. sativa*, and here in *T. aestivum* (Repetto et al., 2008; Choudhary et al., 2009). Several studies have raised the possibility of nuclear translation (Dahlberg et al., 2003), which could take place in the nucleoplasm and the nucleolus in mammalian cells (David et al., 2012). Indeed the idea of a translasome has been described, a super-complex identified in the nucleus of yeast cells which would consist of an assembly of ribosomal proteins, elongation factors, proteasome, chaperones and tRNA synthetases (Sha et al., 2009).

Proteins classified in the nutrient storage activity group were identified in many protein spots. They are very abundant at the end of the grain-filling phase and probably are contaminants. Their presence may even explain the variability in the amount of proteins extracted from grains harvested at 750°Cd after anthesis.

Constant Need for Ribosome Biogenesis Proteins Increasing at the End of Grain Development

Among the 18 ribosomal proteins identified, six were present in multiple spots and for 10 at least one of their protein spots varied in relative abundance during grain development. These variant proteins had abundance profiles 3, 4, 5, or 6. Thus, ribosome formation seems to occur in each phase of grain development. However, more different proteins were present in profiles 4, 5, and 6 (nine different proteins in total) than in the other three profiles (one protein) and were thus more abundant as a class after 450°Cd after anthesis. Possibly there is a high demand for ribosome synthesis during the second half of the effective grain-filling and early maturation phases. This is somewhat similar to *M. truncatula* seed in which transcripts encoding genes involved in ribosome biogenesis were more abundant at the end of seed development (Repetto et al., 2008). This result suggests that there may be a ribosome pool formed at the end of the wheat grain development, which is necessary for grain germination. This is in accordance with results in *Arabidopsis thaliana* which showed that potential for seed germination is largely programmed during the seed maturation phase (Rajjou et al., 2004).

Potential Regulators of the Beginning and End of the Filling Phase

RNA helicases 2 and 34 and Mago nashi-like proteins, all involved in mRNA maturation, were most abundant at 250 and 150°Cd after anthesis, respectively. Similarly, HMG1/2-like protein was most abundant during the cellularization phase. HMG are abundant DNA–binding chromosomal non-histone proteins which are not essential for chromatin organization but act with transcription factors in transcriptional control (Calogero et al., 1999; Jerzmanowski et al., 2000). They are probably also architectural factors in the assembly of certain nucleoprotein complexes (Jerzmanowski et al., 2000). Staining of the 2D gels with Pro-Q Diamond® (Invitrogen) revealed that one spot of this protein was phosphorylated at 150°Cd (Supplementary Figure 3). The histone deacetylase HDAC2 was most abundant at 150°Cd after anthesis. In *A. thaliana* this protein is one of 18 histone deacetylases involved in the repression of gene expression in multiple developmental processes by causing chromatin compaction (Hollender and Liu, 2008; Liu et al., 2014). Staining of the 2DE gels with Pro-Q Diamond® showed that the two spots corresponding to this protein were phosphorylated at 150°Cd after anthesis and one was still phosphorylated at 250°Cd after anthesis (Supplementary Figure 3). In mammals, many HDACs were found to be phosphorylated both *in vitro* and *in vivo* (Sengupta and Seto, 2004). This post-translational modification could affect the activity of this protein. These results suggest that some changes in transcriptional regulation occur at the transition between the cellularization and the grain-filling phases and that a number of genes are repressed in early grain development. In barley grain, a massive transcriptional reprogramming occurs during this developmental transition (Sreenivasulu et al., 2004, 2010) and the proteins identified in the present study may be potential regulators of this key stage.

Identified histones had a higher relative abundance between 150 and 450°Cd after anthesis than later during grain development and more variant types were also detected during this period. This was surprising as some histones have previously been shown to remain constant throughout wheat grain development with the synthesis of histones at 3 dap ending around 16 dap, approximately 320°Cd after anthesis (Spiker et al., 1987).

Several proteins accumulated in the nucleus at the end of grain development. For example, arginine/serine rich splicing factor which is known to be localized in nuclear specks and to be part of the spliceosome (Tillemans et al., 2006). Serpin proteins, which negatively regulate proteases, were also more abundant in late development, as previously observed in the

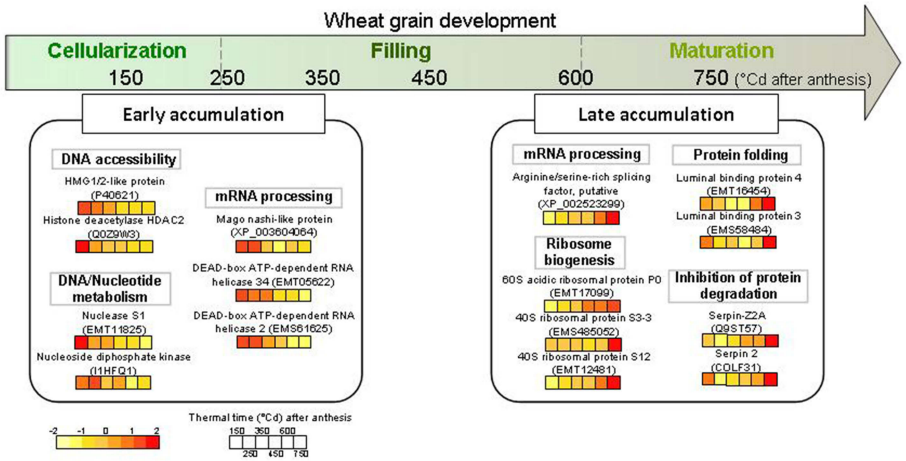


FIGURE 7 | Schematic view of some highlighted nuclear proteins accumulated during early and late grain development. These proteins were extracted from wheat grain at six thermal times after anthesis, from 150 to 750°Cd and analyzed on 2-D gel electrophoresis. The color code for normalized volume of correspondent protein spots is indicated at the bottom left.

endosperm of developing wheat grain (Tasleem-Tahir et al., 2012). Interestingly, proteins involved in proteolysis were more abundant between 150 and 450°Cd after anthesis, before serpins accumulate. Proteins with a role in mRNA maturation or in protection against degradation might regulate processes at the end of the grain-filling phase.

Do Non-variant Nuclear Proteins have a Regulatory Role During Grain Development?

Proteins which did not vary are likely to be essential throughout grain development and may still regulate grain development. An example is the FACT complex subunit SSRP1-B, which like HMG1-2 facilitates the formation of nucleoprotein structures (Röttgers et al., 2000). It may also act in protein complexes to control transcription mechanisms modulating the properties of chromatin. Another example is the histone deacetylase HDT2, which could repress transcription in the same manner as *Z. mays* HDAC2, by forming a complex of three polypeptides (Hollender and Liu, 2008). These proteins probably play an important role during grain development even though we didn't see any change in their abundance. Interestingly, the histone deacetylase HDT2 protein spot was stained with both CBB and Pro-Q Diamond® at 150 and 250°Cd after anthesis, indicating that this protein was phosphorylated at the end of the cellularization phase (Supplementary Figure 3). Perhaps post-translational modification such as phosphorylation regulates these proteins during grain development.

CONCLUDING REMARKS

Some nuclear proteins are central actors in biological processes that regulate seed development. This study identified 114 different wheat proteins with various functions and dynamics, some of which had been found in previous studies of nuclear proteomes of other plant species and organs. For the first time, we have an overview of some of the quantitative changes occurring in 2D nuclear proteome of the developing wheat grain. This study revealed that the dynamics of the nuclear proteome of wheat grain seems to be divided into two periods (**Figure 7**). The first phase corresponds to the cellularization and early effective grain-filling phases, during which a change in transcription regulation occurs with a high abundance of proteins involved in mRNA processing. The second phase corresponds to the end of the effective grain-filling phase and the early maturation phase, when there is an activation of ribosome synthesis and an increase in proteins inhibiting protease action. This study opens the way for more precise research into the regulatory mechanisms that govern the accumulation of starch, storage proteins, and micronutrients that determine the processing and health value of cereal grains.

ACKNOWLEDGMENTS

The authors thank Didier Viala (INRA, Clermont-Ferrand) for help with MS analyses and protein identification, Joelle Henry-Berger and Christelle Damon (Blaise Pascal University, Clermont-Ferrand) for help with microscopic observation, and Dr. Philippe Leroy (INRA, Clermont-Ferrand) for compiling the database for functional annotation of proteins. This work was supported by a Ph.D. grant from the French Ministry for Higher Education and Research to TB and funding from the French Government managed by the Research National Agency (ANR) in the framework of Investments for the Future (ANR-10-BTBR-03), France AgriMer and the French Fund to support Plant Breeding (FSOV).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2015.00905

REFERENCES

Aki, T., and Yanagisawa, S. (2009). Application of rice nuclear proteome analysis to the identification of evolutionarily conserved and glucose-responsive nuclear proteins research articles. *J. Proteome Res.* 8, 3912–3924. doi: 10.1021/pr900187e

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bancel, E., Bonnot, T., Davanture, M., Branlard, G., Zivy, M., and Martre, P. (2015). Proteomic approach to identify nuclear proteins in wheat grain. *J. Proteome Res.* 14, 4432–4439. doi: 10.1021/acs.jproteome.5b00446

Blum, T., Briesemeister, S., and Kohlbacher, O. (2009). MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics* 10:274. doi: 10.1186/1471-2105-10-274

Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., and Lamond, A. I. (2007). The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* 8, 574–585. doi: 10.1038/nrm2184

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Briesemeister, S., Rahnenführer, J., and Kohlbacher, O. (2010). YLoc–an interpretable web server for predicting subcellular localization. *Nucleic Acids Res.* 38, W497–W502. doi: 10.1093/nar/gkq477

Calikowski, T. T., Meulia, T., and Meier, I. (2003). A proteomic study of the arabidopsis nuclear matrix. *J. Cell. Biochem.* 90, 361–378. doi: 10.1002/jcb.10624

Calogero, S., Grassi, F., Aguzzi, A., Voigtländer, T., Ferrier, P., Ferrari, S., et al. (1999). The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nat. Genet.* 22, 276–280. doi: 10.1038/10338

Chojecki, A. J. S., Bayliss, M. W., and Gale, M. D. (1986). Cell production and DNA accumulation in the wheat endosperm, and their association with grain weight. *Ann. Bot.* 58, 809–817.

Choudhary, M. K., Basu, D., Datta, A., Chakraborty, N., and Chakraborty, S. (2009). Dehydration-responsive nuclear proteome of rice (*Oryza sativa* L.) illustrates protein network, novel regulators of cellular adaptation, and evolutionary perspective. *Mol. Cell. Proteomics* 8, 1579–1598. doi: 10.1074/mcp.M800601-MCP200

Dahlberg, J. E., Lund, E., and Goodwin, E. B. (2003). Nuclear translation: what is the evidence? *RNA* 9, 1–8. doi: 10.1261/rna.2121703

David, A., Dolan, B. P., Hickman, H. D., Knowlton, J. J., Clavarino, G., Pierre, P., et al. (2012). Nuclear translation visualized by ribosome-bound nascent chain puromycylation. *J. Cell Biol.* 197, 45–57. doi: 10.1083/jcb.201112145

Erhardt, M., Adamska, I., and Franco, O. L. (2010). Plant nuclear proteomics–inside the cell maestro. *FEBS J.* 277, 3295–3307. doi: 10.1111/j.1742-4658.2010.07748.x

Ferreira, B. S., Branco, A. T., de Oliveira, M. A., Pereira, M. G., and de Souza Filho, G. A. (2006). Methodological improvements on extraction of nuclear proteins and its preliminary analysis during the maize (*Zea mays* L.) endosperm development. *Protein Pept. Lett.* 13, 981–984. doi: 10.2174/092986606778777470

Ferreira, M. S. L., Martre, P., Mangavel, C., Girousse, C., Rosa, N. N., Samson, M.-F., et al. (2012). Physicochemical control of durum wheat grain filling and glutenin polymer assembly under different temperature regimes. *J. Cereal Sci.* 56, 58–66. doi: 10.1016/j.jcs.2011.11.001

Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003). Ribosome assembly in eukaryotes. *Gene* 313, 17–42. doi: 10.1016/S0378-1119(03)00629-2

Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., et al. (2014). LocTree3 prediction of localization. *Nucleic Acids Res.* 42, W350–W355. doi: 10.1093/nar/gku396

Guo, B., Chen, Y., Li, C., Wang, T., Wang, R., Wang, B., et al. (2014). Maize (*Zea mays* L.) seedling leaf nuclear proteome and differentially expressed proteins between a hybrid and its parental lines. *Proteomics* 14, 1071–1087. doi: 10.1002/pmic.201300147

Hollender, C., and Liu, Z. (2008). Histone deacetylase genes in Arabidopsis development. *J. Integr. Plant Biol.* 50, 875–885. doi: 10.1111/j.1744-7909.2008.00704.x

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Hoshikawa, K. (1962). Studies on the development of endosperm in rice. 11. Development of starch granules in endosperm tissue. *Proc. Crop Sci. Soc. Jpn* 37, 207–216. doi: 10.1626/jcs.37.207

Husson, F., Josse, J., Le, S., and Mazet, J. (2014). Package "FactoMineR." Available online at: http://factominer.free.fr

Jaiswal, D. K., Ray, D., Choudhary, M. K., Subba, P., Kumar, A., Verma, J., et al. (2013). Comparative proteomics of dehydration response in the rice nucleus: new insights into the molecular basis of genotype-specific adaptation. *Proteomics* 13, 3478–3497. doi: 10.1002/pmic.201300284

Jerzmanowski, A., Przewłoka, M., and Grasser, K. D. (2000). Linker Histones and HMG1 proteins of higher plants. *Plant Biol.* 2, 586–597. doi: 10.1055/s-2000-16648

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi: 10.1093/nar/gkt1076

Khan, M. M. K., and Komatsu, S. (2004). Rice proteomics: recent developments and analysis of nuclear proteins. *Phytochemistry* 65, 1671–1681. doi: 10.1016/j.phytochem.2004.04.012

Li, G., Nallamilli, B. R. R., Tan, F., and Peng, Z. (2008). Removal of high-abundance proteins for nuclear subproteome studies in rice (*Oryza sativa*) endosperm. *Electrophoresis* 29, 604–617. doi: 10.1002/elps.200700412

Liu, X., Yang, S., Zhao, M., Luo, M., Yu, C.-W., Chen, C.-Y., et al. (2014). Transcriptional repression by histone deacetylases in plants. *Mol. Plant* 7, 764–772. doi: 10.1093/mp/ssu033

Mares, D., Norstog, K., and Stone, B. (1975). Early stages in the development of wheat endosperm. I. The change from free nuclear to cellular endosperm. *Aust. J. Bot.* 23, 311–326. doi: 10.1071/BT9750311

Neuhoff, V., Stamm, R., and Eibl, H. (1985). Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: a systematic analysis. *Electrophoresis* 6, 427–448. doi: 10.1002/elps.1150060905

Nollen, E. A. A., Salomons, F. A., Brunsting, J. F., van der Want, J. J. L., Sibon, O. C. M., and Kampinga, H. H. (2001). Dynamic changes in the localization of thermally unfolded nuclear proteins associated with chaperone-dependent protection. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12038–12043. doi: 10.1073/pnas.201112398

Pandey, A., Choudhary, M. K., Bhushan, D., Chattopadhyay, A., Chakraborty, S., Datta, A., et al. (2006). The nuclear proteome of chickpea (*Cicer arietinum* L.) reveals predicted and unexpected proteins. *J. Proteome Res.* 5, 3301–3311. doi: 10.1021/pr060147a

Petrovská, B., Jeřábková, H., Chamrád, I., Vrána, J., Lenobel, R., Uřinovská, J., et al. (2014). Proteomic analysis of barley cell nuclei purified by flow sorting. *Cytogenet. Genome Res.* 143, 78–86. doi: 10.1159/000365311

Petrovská, B., Šebela, M., and Doležel, J. (2015). Inside a plant nucleus: discovering the proteins. *J. Exp. Bot.* 66, 1627–1640. doi: 10.1093/jxb/erv041

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The Effect of α-Amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* 134, 1598–1613. doi: 10.1104/pp.103.036293

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Repetto, O., Rogniaux, H., Firnhaber, C., Zuber, H., Küster, H., Larré, C., et al. (2008). Exploring the nuclear proteome of *Medicago truncatula* at the switch towards seed filling. *Plant J.* 56, 398–410. doi: 10.1111/j.1365-313X.2008.03610.x

Repetto, O., Rogniaux, H., Larré, C., Thompson, R., and Gallardo, K. (2012). The seed nuclear proteome. *Front. Plant Sci.* 3:289. doi: 10.3389/fpls.2012.00289

Romeuf, I., Tessier, D., Dardevet, M., Branlard, G., Charmet, G., and Ravel, C. (2010). wDBTF: an integrated database resource for studying wheat transcription factor families. *BMC Genomics* 11:185. doi: 10.1186/1471-2164-11-185

Röttgers, K., Krohn, N. M., Lichota, J., Stemmer, C., Merkle, T., and Grasser, K. D. (2000). DNA-interactions and nuclear localisation of the chromosomal HMG domain protein SSRP1 from maize. *Plant J.* 23, 395–405. doi: 10.1046/j.1365-313x.2000.00801.x

Sabelli, P. A, and Larkins, B. A (2009). The development of endosperm in grasses. *Plant Physiol.* 149, 14–26. doi: 10.1104/pp.108.129437

Schnyder, H., and Baum, U. (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. doi: 10.1016/S1161-0301(14)80001-4

Sengupta, N., and Seto, E. (2004). Regulation of histone deacetylase activities. *J. Cell. Biochem.* 93, 57–67. doi: 10.1002/jcb.20179

Sha, Z., Brill, L. M., Cabrera, R., Kleifeld, O., Scheliga, J. S., Glickman, M. H., et al. a (2009). The eIF3 interactome reveals the translasome, a supercomplex linking protein synthesis and degradation machineries. *Mol. Cell* 36, 141–152. doi: 10.1016/j.molcel.2009.09.026

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* 68, 850–858. doi: 10.1021/ac950914h

Shewry, P. R., Mitchell, R. A. C., Tosi, P., Wan, Y., Underwood, C., Lovegrove, A., et al. (2012). An integrated study of grain development of wheat (cv. Hereward). *J. Cereal Sci.* 56, 21–30. doi: 10.1016/j.jcs.2011.11.007

Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., et al. (2001). The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J. Biol. Chem.* 276, 33293–33296. doi: 10.1074/jbc.R100016200

Spiker, S., Hopkins, R., Fischer, R., and Quatrano, R. S. (1987). Synthesis of nucleosomal histone variants during wheat grain development. *Biochim. Biophys. Acta Gene Struct. Expr.* 910, 157–162. doi: 10.1016/0167-4781(87)90068-6

Sreenivasulu, N., Altschmied, L., Radchuk, V., Gubatz, S., Wobus, U., and Weschke, W. (2004). Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. *Plant J.* 37, 539–553. doi: 10.1046/j.1365-313X.2003.01981.x

Sreenivasulu, N., Borisjuk, L., Junker, B. H., Mock, H.-P., Rolletschek, H., Seiffert, U., et al. (2010). Barley grain development toward an integrative view. *Int. Rev. Cell Mol. Biol.* 281, 49–89. doi: 10.1016/S1937-6448(10)81002-0

Tan, F., Li, G., Chitteti, B. R., and Peng, Z. (2007). Proteome and phosphoproteome analysis of chromatin associated proteins in rice (*Oryza sativa*). *Proteomics* 7, 4511–4527. doi: 10.1002/pmic.200700580

Tasleem-Tahir, A., Nadaud, I., Chambon, C., and Branlard, G. (2012). Expression profiling of starchy endosperm metabolic proteins at 21 stages of wheat grain development. *J. Proteome Res.* 11, 2754–2773. doi: 10.1021/pr201110d

Tillemans, V., Leponce, I., Rausin, G., Dispa, L., and Motte, P. (2006). Insights into nuclear organization in plants as revealed by the dynamic distribution of Arabidopsis SR splicing factors. *Plant Cell* 18, 3218–3234. doi: 10.1105/tpc.106.044529

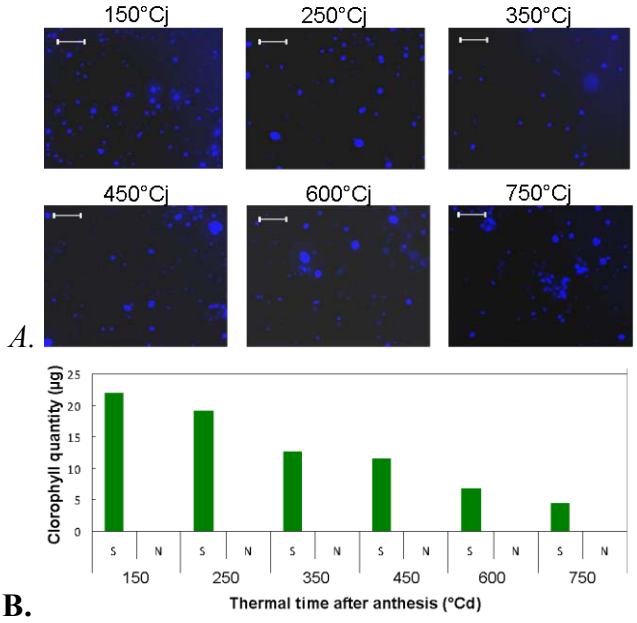
Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 32, 223–226. doi: 10.1038/nbt.2839

Willard, F. S., and Crouch, M. F. (2000). Nuclear and cytoskeletal translocation and localization of heterotrimeric G-proteins. *Immunol. Cell Biol.* 78, 387–394. doi: 10.1046/j.1440-1711.2000.00927.x

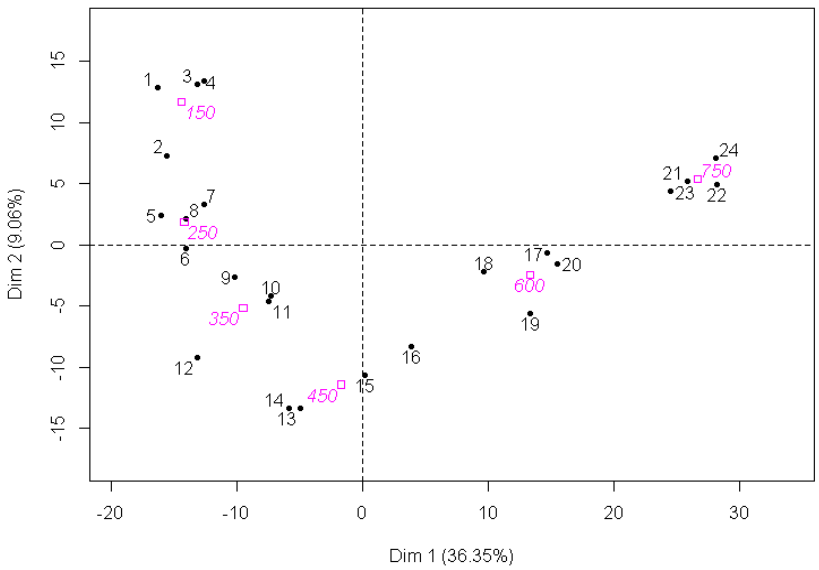
Zhang, S., Xia, X., Shen, J., Zhou, Y., and Sun, Z. (2008). DBMLoc: a Database of proteins with multiple subcellular localizations. *BMC Bioinformatics* 9:127. doi: 10.1186/1471-2105-9-127

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

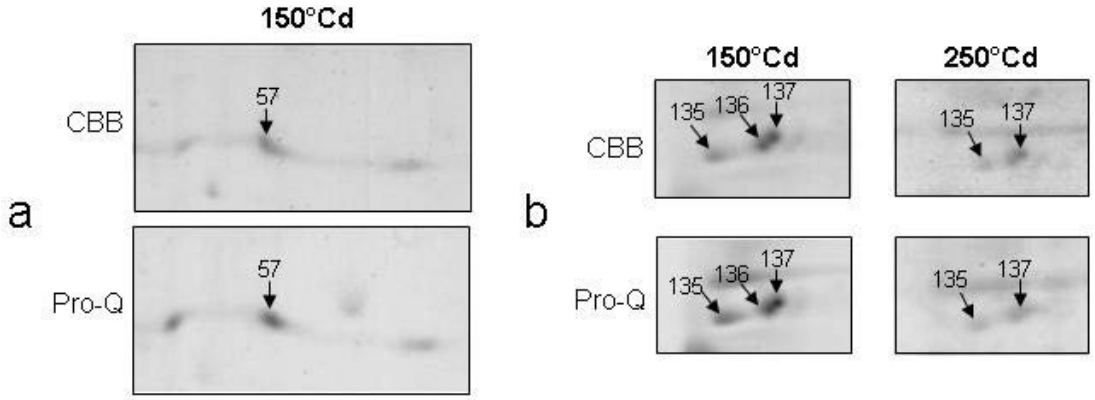
Supplementary information



Supplementary Figure 1. Validation of nuclei purification. Nuclei were isolated from wheat grains harvested between 150°C and 750°Cd after anthesis. A. Nuclei were observed under fluorescence microscopy (Zeiss Axioplan 2 microscope). Scale bar represents 50 µm sizes.
B. Chlorophyll assay was performed on supernatants (S) collected during washing steps of nuclei and on nuclei extracts (N).



Supplementary Figure 2. Principal component analysis of protein spots detected by image analysis. Normalized volumes of protein spots detected in the four replicates were collected. 1-4: replicates at 150°Cd; 5-8: replicates at 250°Cd; 9-12: replicates at 350°Cd; 13-16: replicates at 450°Cd; 17-20: replicates at 600°Cd; 21-24: replicates at 750°Cd.



Supplementary Figure 3. Some nuclear protein spots phosphorylated during division / differentiation phase of wheat grain development. 2D gels were stained with Pro-Q Diamond (Pro-Q, Invitrogen) according to the manufacturer's instructions and adapted (Agrawal and Thelen, 2005) then stained with Coomassie Brilliant Blue G250 (CBB). Proteins spots stained both with Pro-Q and CBB were considered as phosphorylated. **a**: HMG1/2-like protein (spot 57) was phosphorylated at 150°Cd after anthesis. **b**: Histone deacetylase HDAC2 was phosphorylated at 150°Cd (spot 135, 136) and 250°Cd after anthesis (spot 135); Histone deacetylase HDT2 was phosphorylated at 150 and 250°Cd (spot 137) after anthesis.

Reference:

Agrawal, G. K., and Thelen, J. J. (2005). Development of a simplified, economical polyacrylamide gel staining protocol for phosphoproteins. *Proteomics* 5, 4684–8. doi:10.1002/pmic.200500021.

Supplementary table 1 and table 2 can be viewed online at:
http://journal.frontiersin.org/article/10.3389/fpls.2015.00905/full

3. CONCLUSIONS

Cette étude a permis pour la première fois d'analyser les variations d'abondance des protéines nucléaires au cours du développement du grain de blé. D'un point de vue qualitatif, les fonctions représentées au sein des 114 protéines identifiées sont comparables à celles retrouvées dans le chapitre précédent. En effet, nous avons retrouvé de nombreuses protéines ribosomales ainsi que plusieurs protéines histones, ce qui était attendu. Plusieurs protéines liées à la conformation de l'ADN, à la transcription et sa régulation, ou à la maturation des ARNm ont également été identifiées.

La principale difficulté rencontrée dans cette étude a été la présence de protéines de réserve en quantité non négligeable dans les extraits nucléaires, représentant ainsi des contaminants. Du fait de leur présence sous plusieurs isoformes, ces protéines ont été identifiées dans de nombreux spots protéiques prélevés sur les gels bidimensionnels. Ces spots étaient particulièrement abondants à 600 et 750°Cj après floraison. Les protéines de réserve, en très grande abondance dans le grain en fin de phase de remplissage posent donc un problème pour l'analyse du protéome nucléaire. L'hypothèse que nous émettons est que les corpuscules protéiques dans lesquels s'accumulent les protéines de réserve du grain ont pu être co-purifiés avec les noyaux sur le gradient de densité car ils auraient à ces stades une densité proche de celle des noyaux cellulaires. De plus, à cette période du développement du grain les noyaux sont probablement plus difficilement purifiables du fait de leur dégradation progressive dans la mort cellulaire programmée, orchestrée dans l'albumen au cours de la phase de maturation (Ferreira *et al.*, 2012 ; Dominguez and Cejudo, 2014). Nous avons ainsi fait le choix pour les prochaines analyses du protéome nucléaire de n'extraire les noyaux que pour des stades antérieurs à 600°Cj après floraison.

Malgré cette difficulté, nous avons pu voir que les quantités de protéines nucléaires extraites étaient comparables pour un même nombre de grains entre 150 et 600°Cj après floraison. L'analyse des images de gels et les analyses statistiques ont démontré qu'une grande majorité des spots protéiques révélés sur les gels d'électrophorèse a présenté des variations d'abondance sur la cinétique considérée. Le stade de développement du grain a donc un effet très important sur la dynamique du protéome nucléaire. Au vu de leur profil d'accumulation, certaines protéines liées à la conformation de l'ADN (protéine HMG, histone déacétylase) pourraient jouer un rôle important dans la transition entre la phase de cellularisation et la phase de remplissage.

Les fortes variations quantitatives observées au niveau des protéines nucléaires pendant le développement du grain nous ont ainsi laissé entrevoir la possibilité d'étudier la réponse de ce sous-protéome à d'autres facteurs tels que la nutrition de la plante.

Références :

Dominguez, F. and Cejudo, F.J. (2014) Programmed cell death (PCD): an essential process of cereal seed development and germination. *Front. Plant Sci.*, **5**, 366.

Ferreira, M.S.L., Martre, P., Mangavel, C., Girousse, C., Rosa, N.N., Samson, M.-F. and Morel, M.-H. (2012) Physicochemical control of durum wheat grain filling and glutenin polymer assembly under different temperature regimes. *J. Cereal Sci.*, **56**, 58–66.

CHAPITRE 4 :
La réponse protéomique du grain de *Triticum monococcum* à la nutrition azotée et soufrée

1. INTRODUCTION

Les effets de l'N et du S sur la quantité totale de PR et sur la proportion des différentes classes et sous-unités de PR du grain de blé ont largement été démontrés. En revanche, les mécanismes moléculaires qui interviennent entre la nutrition du grain et l'ajustement de la synthèse des PR ont été peu caractérisés. Les approches -omiques nous sont apparues comme les plus adaptées pour apporter de nouveaux éléments à la compréhension de ces mécanismes. Notre hypothèse de travail a été la suivante : les quantités d'N et de S disponibles pour le grain lors du remplissage affectent l'abondance de nombreux transcrits, protéines et métabolites interconnectés, qui agiraient en amont de l'expression des gènes de PR.

Afin de visualiser la réponse du grain à la nutrition azotée et soufrée, nous avons conduit une expérimentation chez *Triticum monococcum*. Cette espèce de blé est un bon modèle d'étude du fait de son génome diploïde, plus simple que le génome hexaploïde du blé tendre et en même temps très proche de son sous-génome A. De plus, les PR présentes chez cette espèce sont les mêmes que celles du blé tendre. Dans cette expérimentation, quatre nutritions contenant des quantités différentes en N et S ont été appliquées à la culture à partir de 200°Cj après floraison, c'est-à-dire en fin de phase de cellularisation du grain (Figure 23). La première nutrition ne contenait ni apport d'N ni de S et a été appelée N-S-, la seconde, nommée N+S-, correspondait à une nutrition azotée dépourvue de soufre (6mM N, 0mM S), la troisième, N-S+, était une nutrition soufrée (0,5mM N, 2mM S) et la quatrième nutrition, N+S+, combinait des fortes doses des deux éléments (6mM N, 2mM S).

Les grains ont été récoltés tous les 100°Cj après floraison dans le but de réaliser différentes analyses omiques (protéomique, transcriptomique, métabolomique). Nous avons fait le choix de traiter ces données dans deux études différentes :
- La première étude correspond à la réponse protéomique du grain à la nutrition et sera décrite dans ce chapitre 4.
- La seconde correspond à l'analyse du transcriptome et des métabolites. Afin d'obtenir une vision intégrative de la réponse du grain à la nutrition, ces données ont été intégrées aux données de protéomique obtenues dans le chapitre 4. Ce travail est traité dans le chapitre 5.

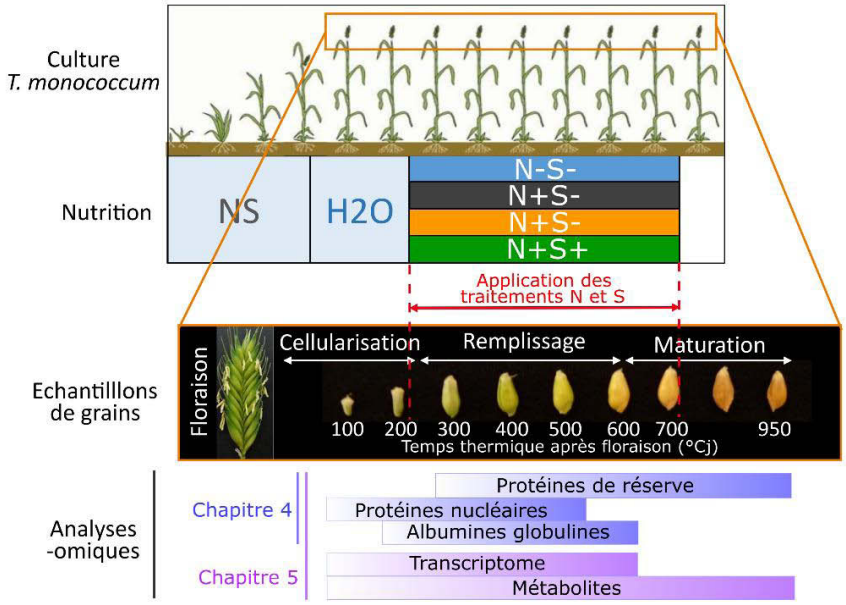


Figure 23. Schéma expérimental de l'étude de la réponse du grain de Triticum monococcum à la nutrition azotée et soufrée.

Dans ce chapitre 4 qui concerne la réponse protéomique du grain à la nutrition azotée et soufrée, nous avons ciblé différents sous-protéomes : les PR, les albumines globulines qui correspondent majoritairement à des enzymes ou inhibiteurs d'enzymes, et les protéines nucléaires. Nous avons ainsi voulu visualiser suite à une modification de la nutrition de la plante les changements qui s'opèrent au niveau du métabolisme cellulaire et des régulateurs transcriptionnels, changements qui pourraient influencer la synthèse des PR. Les trois sous-protéomes ont tous été analysés entre 300 et 500°Cj, c'est-à-dire pendant le remplissage du grain et après l'application des quatre traitements. Les PR ont elles été quantifiées jusqu'à maturité du grain pour suivre leurs profils d'accumulation et observer la composition du grain mature. D'autres stades ont été analysés pour obtenir un maximum d'information sur l'identification et la quantification des protéines (100 et 200°Cj après floraison pour les protéines nucléaires, 200 et 600°Cj après floraison pour les albumines globulines). Comme nous l'avons évoqué, nous avons fait le choix de ne pas extraire les protéines nucléaires des grains au-delà de 500°Cj pour des raisons techniques (cf conclusion chapitre 3). Pour identifier et quantifier les protéines nucléaires et albumines globulines, nous avons opté pour une analyse des échantillons protéiques en spectrométrie de masse, sans passer au préalable par une étape de séparation sur gel d'électrophorèse (gel free).

2. ARTICLE 3 : Grain subproteome responses to nitrogen and sulfur supply in diploid wheat *Triticum monococcum* ssp. *monococcum*

Titouan Bonnot[1], Emmanuelle Bancel[1], David Alvarez[1], Marlène Davanture[2], Julie Boudet[1], Marie Pailloux[3], Michel Zivy[2], Catherine Ravel[1] and Pierre Martre[1]

[1] *UMR GDEC, INRA, Université Blaise Pascal, 5 chemin de Beaulieu, 63 177 Aubière, France*
[2] *UMR GQE, INRA, Université Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91 190 Gif-sur-Yvette, France*
[3] *LIMOS, CNRS, Université Blaise Pascal, 63 173 Aubière, France*
[†]*Present address: UMR LEPSE, INRA, Montpellier SupAgro, 34 060 Montpellier, France*

(Article soumis à *Plant Journal*)

Grain subproteome responses to nitrogen and sulfur supply in diploid wheat *Triticum monococcum* ssp. *monococcum*

Titouan Bonnot[1], Emmanuelle Bancel[1], David Alvarez[1], Marlène Davanture[2], Julie Boudet[1], Marie Pailloux[3], Michel Zivy[2], Catherine Ravel[1,*] and Pierre Martre[1,†]

[1] UMR GDEC, INRA, Université Blaise Pascal, 5 chemin de Beaulieu, 63 177 Aubière, France

[3] UMR GQE, INRA, Université Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91 190 Gif-sur-Yvette, France

[4] LIMOS, CNRS, Université Blaise Pascal, 63 173 Aubière, France

[†]Present address: UMR LEPSE, INRA, Montpellier SupAgro, 34060 Montpellier, France

SUMMARY

Wheat grain storage proteins (GSPs) make up most of the protein content of grain and determine flour end-use value. The synthesis and accumulation of GSPs depend highly on nitrogen (N) and sulfur (S) availability and it is important to understand the underlying control mechanisms. Here we studied how the einkorn (*Triticum monococcum* ssp. *monococcum*) grain proteome responds to different amounts of N and S supply during grain development. GSP composition at grain maturity was clearly impacted by nutrition treatments, due to early changes in the rate of GSP accumulation during grain filling. Large-scale analysis of the nuclear and albumin-globulin sub-proteomes during this key developmental phase revealed that the abundance of 203 proteins was significantly modified by the nutrition treatments. Our results showed that the grain proteome was highly impacted by perturbation in the N:S balance. S supply strongly increased the rate of accumulation of S-rich α/β-gliadin and γ-gliadin, and of several other proteins involved in glutathione metabolism. Post-anthesis N supply resulted in the activation of amino acid metabolism at the expense of carbohydrate metabolism and the activation of transport processes including nucleocytoplasmic transit. Protein accumulation networks were analyzed. Several central actors in the response were identified whose variation in abundance was related to variation in the amounts of many other proteins and are thus potentially important for GSP accumulation. This detailed analysis of grain sub-proteomes provides information on how wheat GSP composition can possibly be controlled in low-level fertilization condition.

Keywords: einkorn, grain, nitrogen, sulfur, storage proteins, nuclear proteins, albumin-globulin, network.

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is the main stable crop in many regions of the world and globally provides about 20% of calories and proteins in the human diet. In mature grain, 10-15% of the dry mass is protein. GSPs make up 60-80% of the total protein in grain and metabolic proteins (the albumins and globulins) 15-20% (Shewry and Halford, 2002). GSPs accumulate during the effective filling phase of grain development (Shewry *et al.*, 2012). The two main GSP fractions are glutenin and gliadin that when mixed together with water form gluten and are thus the main determinants of the rheological and bread-making properties of wheat dough. More precisely, glutenins form polymers that are mainly responsible for the viscoelastic properties of dough and gliadins are monomeric proteins which confer extensibility (Branlard *et al.*, 2001). The GSP composition of wheat grain thus determines its use-end value. Commonly, gliadins are subdivided into ω1,2-, ω5-, α/β- and γ-gliadin classes and glutenins are classified as high-molecular-weight subunits (HMW-GS) or low-molecular-weight subunits (LMW-GS; Wieser, 2007). These different GSP subclasses and subunits differ in the proportions of cysteine and methionine residues they contain, and are thus classified as being S-rich (α/β- and γ-gliadins, LMW-GS) or S-poor (ω1,2- and ω5-gliadins, HMW-GS; Shewry *et al.*, 1997; Shewry *et al.*, 2001).

The availability of N and S in the soil highly influences the GSP composition of grain. N supply increases the rate and duration of protein

accumulation and so increases the proportion of S-poor GSPs in mature grain (Shewry *et al.*, 2001; Triboï *et al.*, 2003; Chope *et al.*, 2014). In S-deficient conditions, GSP synthesis and accumulation occurs earlier because the cellularization phase is shortened (Castle and Randall, 1987). Mature grain contains low concentrations of cysteine and methionine when S is deficient (Wrigley *et al.*, 1980), causing S-rich GSPs to accumulate more slowly, which is compensated by an increased rate of accumulation of S-poor GSPs, particularly HMW-GS (Zhao *et al.*, 1999; Wieser *et al.*, 2004; Zörb *et al.*, 2010). The molecular mechanisms that control GSP synthesis and composition in response to N and S supply are not yet well understood (Dai *et al.*, 2015). A priority for wheat breeders is to improve grain yield while reducing the need for fertilizers, especially N input. There is a strongly negative correlation between grain yield and grain protein concentration traits, so increasing grain yield is generally detrimental to grain protein concentration and hence grain quality (Simmonds, 1995; Oury and Godin, 2007). In addition, S deficiency is more frequently observed in soils nowadays, partly because of reductions in industrial SO_2 emissions (Eriksen, 2009), hindering the synthesis of S-rich GSPs in wheat. Analyzing responses to N and S nutrition will reveal key actors that can be targeted in developing plants that produce high quality grain in low-level fertilization conditions.

Large-scale proteomics and other omics approaches are commonly used to gain an overview of changes taking place in an organism or tissue by identifying the proteins and genes involved in a given biological response (Das *et al.*, 2015). When proteomics is used to study the grain response to nutrition, bioactive proteins are expected to be found either in the albumin and globulin fraction or among regulatory proteins. Albumins and globulins are mainly enzymes or enzyme inhibitors involved in cell metabolism and development and they may thus influence the rheological properties of wheat flour (Hill *et al.*, 2008). Albumin-globulin synthesis depends on the nutritional status of the plant. Although the synthesis of such proteins is less affected by N than that of GSPs (Wieser and Seilmeier, 1998; Martre *et al.*, 2003), N and S nutrition has been reported to affect several enzymes in wheat grain. For example, a glyceraldehyde-3-phosphate dehydrogenase and a serpin were both increased in conditions combining high N with low S (Flæte *et al.*, 2005). As GSP synthesis is controlled at a transcriptional level and many regulatory factors are nuclear proteins, quality determinants were expected to be nuclear localized (Verdier and Thompson, 2008). The transcriptional regulation of GSP synthesis was confirmed by Dai *et al.* (2015) who reported significantly linear relationships between regulatory gene expression and GSP accumulation rates. Several nuclear proteome studies in other plant species have identified protein actors central to various stress responses (Yin and Komatsu, 2015; Petrovská *et al.*, 2015). Albumin-globulin and nuclear proteins are thus two subproteomes that can be targeted as being involved in GSP synthesis affected by N and S nutrition.

Einkorn (*Triticum monococcum*) is well tolerant of low input cropping and is appreciated for its excellent nutritional properties due to its high protein, high carotenoid and high tocol contents (Corbellini *et al.*, 1999; Hidalgo *et al.*, 2006; Hidalgo and Brandolini, 2014). It is an ancestral wheat whose diploid genome is a sister to bread wheat (*Triticum aestivum*) genome A (Marcussen *et al.*, 2014). Genomic resources are now available for this species. Its genomic sequence is available at https://urgi.versailles.inra.fr/download/iwgsc/TGAC_WGS_assemblies_of_other_wheat_species/ and its transcriptome was analyzed and annotated by Fox *et al.* (2014). Therefore, einkorn appears to be a suitable species in which to explore proteome responses to N and S nutrition.

Our eventual aim is to identify key factors able to ensure high quality grain (high protein content, adequate GSP composition) under low fertilization levels. In this study we produced an overview of the *T. monococcum* grain proteome response to N and S nutrition. Four post-anthesis N and S treatments were applied to a *T. monococcum* genotype grown in a glasshouse and grains were harvested at different developmental stages. GSPs were quantified and N and/or S supply was found to have significant effects on the amounts of gliadin classes and glutenin subunits. In mature grains, α/β- and γ-gliadins were increased by S supply, resulting in an increased gliadin-to-glutenin ratio, while HMW-GS and ω1,2-gliadins were increased with N supply. By focusing on nuclear and albumin-globulin fractions during the effective grain filling phase, 203 proteins were highlighted as being significantly affected by nutrition. Protein data integration grouped proteins with similar accumulation behaviours, which may potentially act synergistically in the grain. A few of these nuclear and albumin-globulin proteins are described in more detail as they potentially play a central role in GSP synthesis control in response to N and S nutrition.

RESULTS

Grain storage protein composition is modified by N and S supply

To study the response of nuclear, albumin-globulin and storage protein subproteomes of developing einkorn wheat grains to N and S supply, four treatments were applied to plants during the effective grain filling period. Plants were grown in a greenhouse in conditions resembling those in the field. The four treatments were N-S-, N+S-, N-S+, N+S+ according to whether N and/or S were applied. Dry mass and protein, N and S content of grain harvested from ears every 100°Cd were measured (Figure 1). Temporal data were fitted with logistic function equations (eqn. 1) to illustrate the rate and duration of compound accumulation (Figure 1 b, d, f) which determine the total quantity of storage compounds at maturity.

The treatments altering N and S availability strongly modified the duration during which N and S accumulated in grains, but the rates of accumulation were much less modified (Figure 1b,d; Table S2). Mature grain contained more N following high N treatment (N+S- and N+S+) compared with the low N and low S treatment (N-S-), mainly because N accumulation lasted longer when more N was supplied (Figure 1b; Tables S1, S2). Similarly, the higher quantity of S per grain at maturity for the high S treatments (N-S+ and N+S+) compared with the low S treatments (N+S- and N-S-) was mainly explained by the longer duration of accumulation of grain S with high S supply (Figure 1d; Table S2). The duration of dry mass accumulation increased when more N and S was supplied, but lower rate of accumulation compensated for this, so the final quantity of dry mass per grain at maturity was not modified by the treatments (Figure 1f; Table S2). Therefore, at maturity concentrations of grain protein and S differed between the treatments and reflected the changes in the quantity of N and S per grain (Figure 1a, c, e; Table S1). In N+S- grain, the N-to-S ratio was higher than 30, an indication of severe S deficiency (Figure 1e).

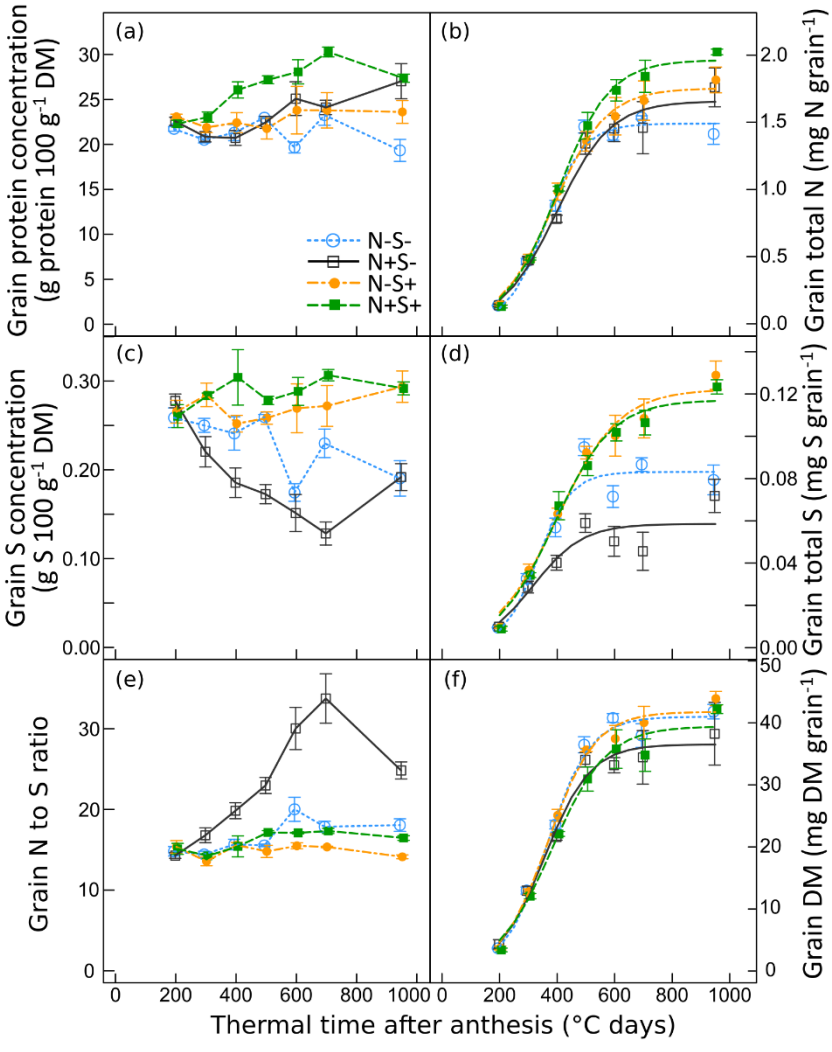


Figure 1. Effects of N and S supply on the accumulation of total grain N, S and dry mass (DM) in einkorn. Total protein concentration (a), mass of N per grain (b), S concentration (c), mass of S per grain (d), N-to-S ratio (e) and grain dry ass (f) were measured. The four treatments N-S-, N+S-, N-S+ and N+S+ were applied from 200 to 700 °Cd after anthesis. Data are means ± 1 s.e. for n = 4 independent replicates. Lines in (b), (d) and (f) were fitted to the data using eqn. (1).

The treatments also had significant effects on the rate and duration of GSP accumulation (Figure 2, Table S2) and the final quantity in mature grains (Figure S1). At maturity, there was 67% more α/β- and 163% more γ-gliadin per grain for the high S treatments (N-S+ and N+S+) compared with the low S treatments (N+S- and N-S-). These S-rich proteins accumulated at higher rates (+29% and +53%) and for longer durations (+30% and +13%) under the higher S supply (Figure 2c, e; Table S2). The quantity per grain of the S-poor ω1,2-gliadin was 45% higher for N+S+ compared with N+S- because accumulation lasted longer (+242°Cd, ~ +11.7 d), while the rate of accumulation was similar (Figure 2d; Table S2). For S-poor HMW-GS, there was 52% more per grain for N+S- than for the other treatments because of a higher rate of accumulation (+106%), which more than compensated for the shorter duration of accumulation (-21%) for N+S- (Figure 2b; Table S2). No significant difference between the treatments was measured at grain maturity for the S-poor ω5-gliadin and the S-rich LMW-GS. For LMW-GS, the effects of N and S on their rate of accumulation was compensated by opposing effects on the duration of their accumulation, which could explain this result (Figure 2a; Table S2). Therefore, except for ω5-gliadin and LMW-GS, the variations in S-rich and S-poor GSP due to the different treatments reflected grain N-to-S ratios at maturity. The gliadin-to-glutenin ratio was different between treatments and increased when grain N-to-S ratio decreased (Figure S1), modifying the GSP composition at maturity (Figure S1d).

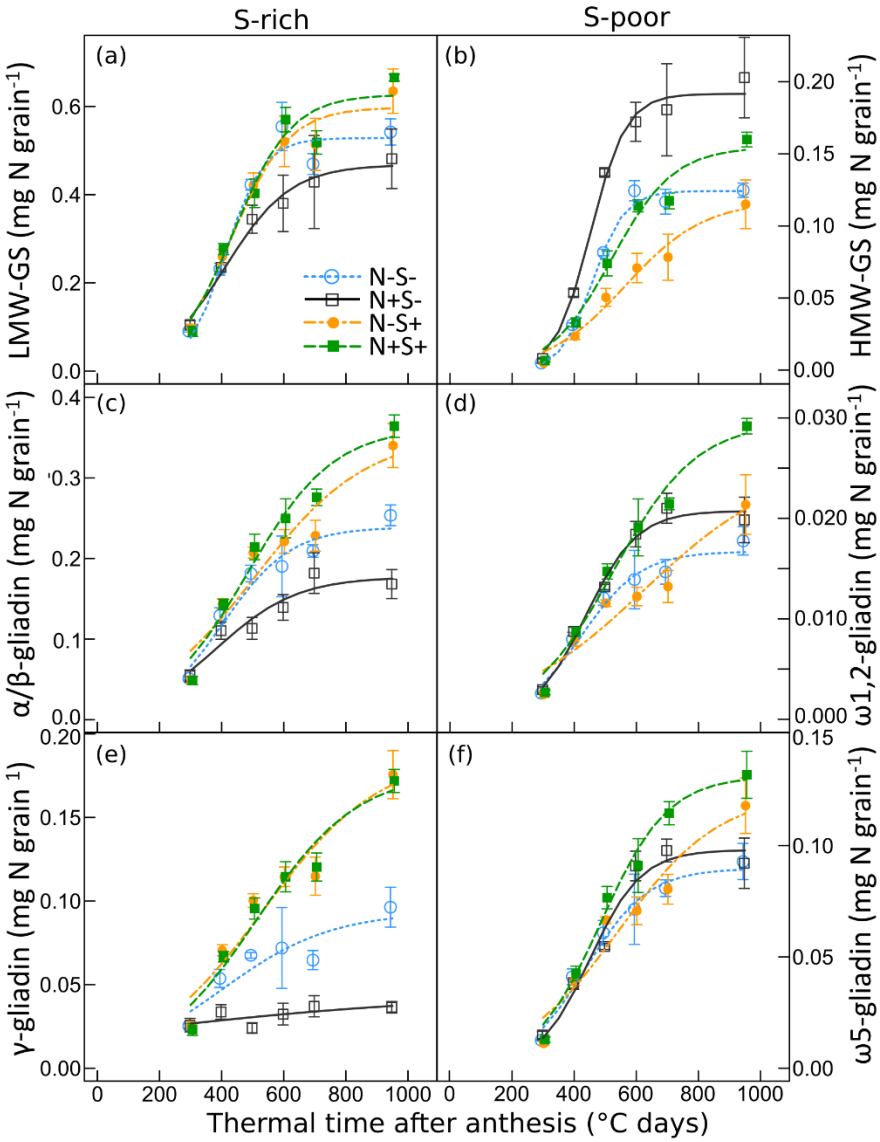


Figure 2. Effects of N and S supply on the accumulation of LMW-GS and HMW-GS and α/β-, γ-, ω5- and ω1,2-gliadin in einkorn grain. The four treatments N-S-, N+S-, N-S+ and N+S+ were applied from 200 to 700 °Cd after anthesis. Data are means ± 1 s.e. for $n = 4$ independent replicates. Lines were fitted to the data using eqn. (1).

Analyzing nuclear and albumin-globulin protein fractions gives high coverage of the grain proteome

Some of the mechanisms by which N and S supply influence GSP accumulation during grain filling must be governed by proteins. To investigate the grain proteome response to nutrition, two subproteomes from the same biological material were analyzed by shotgun mass spectrometry: nuclear proteins from 100 to 500°Cd after anthesis and albumin-globulin proteins from 200 to 600°Cd after anthesis. In total 1677 and 2475 proteins were identified in nuclear and albumin-globulin protein extracts, respectively. These proteins were classified as participating in 24 biological processes and 20 molecular functions according to their gene ontology (GO) information. A total of 978 (24%) different proteins were accurately quantified (466 nuclear proteins and 512 albumin-globulin proteins; Figure 3a, b). Overall 47% and 33% of all quantified nuclear proteins and albumin-globulin respectively were classified as "uncharacterized" as they could not be assigned to a particular biological process. In the albumin-globulin fraction, several proteins were involved in responses to stimulus (12%), protein metabolic processes (11%) and carbohydrate metabolic processes (8%). In nuclear extracts, many proteins were either ribosomal proteins or histones and were thus classified as having functions in protein metabolic processes (17%) and cellular component organization (10%), respectively. Several proteins involved in transport processes were also quantified (5% of the proteins from nuclear and 6% from albumin-globulin subproteomes). A clear difference in GO term enrichment for molecular functions was found between albumin-globulin and nuclear proteins. Many albumin-globulin proteins possess catalytic activity (20%), hydrolase activity (9%) or nucleotide binding activity (12%). In the nuclear extract, an enrichment of proteins with DNA binding (10%), nucleic acid binding (13%) and nucleotide binding (15%) activities was observed as expected. These results highlighted the value of independently analyzing the two subproteomes from the same biological material as the information obtained was not redundant.

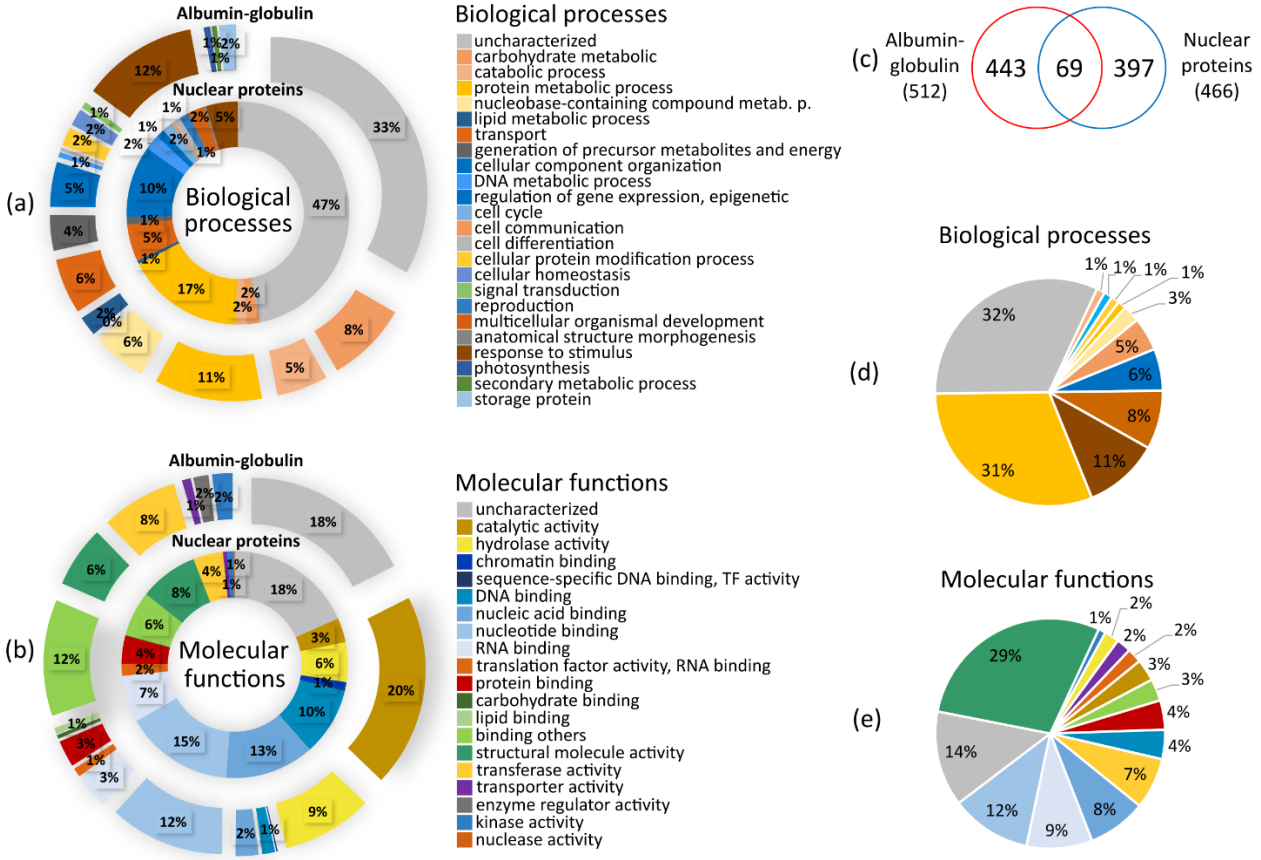


Figure 3. Functional classification of quantified nuclear and albumin-globulin proteins in einkorn grain. Proteins were grouped as being involved in 24 biological processes (a) and 20 molecular functions (b) according to gene ontology annotation. In (a) and (b) nuclear proteins and albumin-globulin are represented in the inner and outer circles, respectively. (c) Venn diagram showing the numbers of quantified proteins in nuclear and albumin-globulin extracts. Biological processes (d) and molecular functions (e) of the 69 proteins found in both nuclear and albumin-globulin protein fractions.

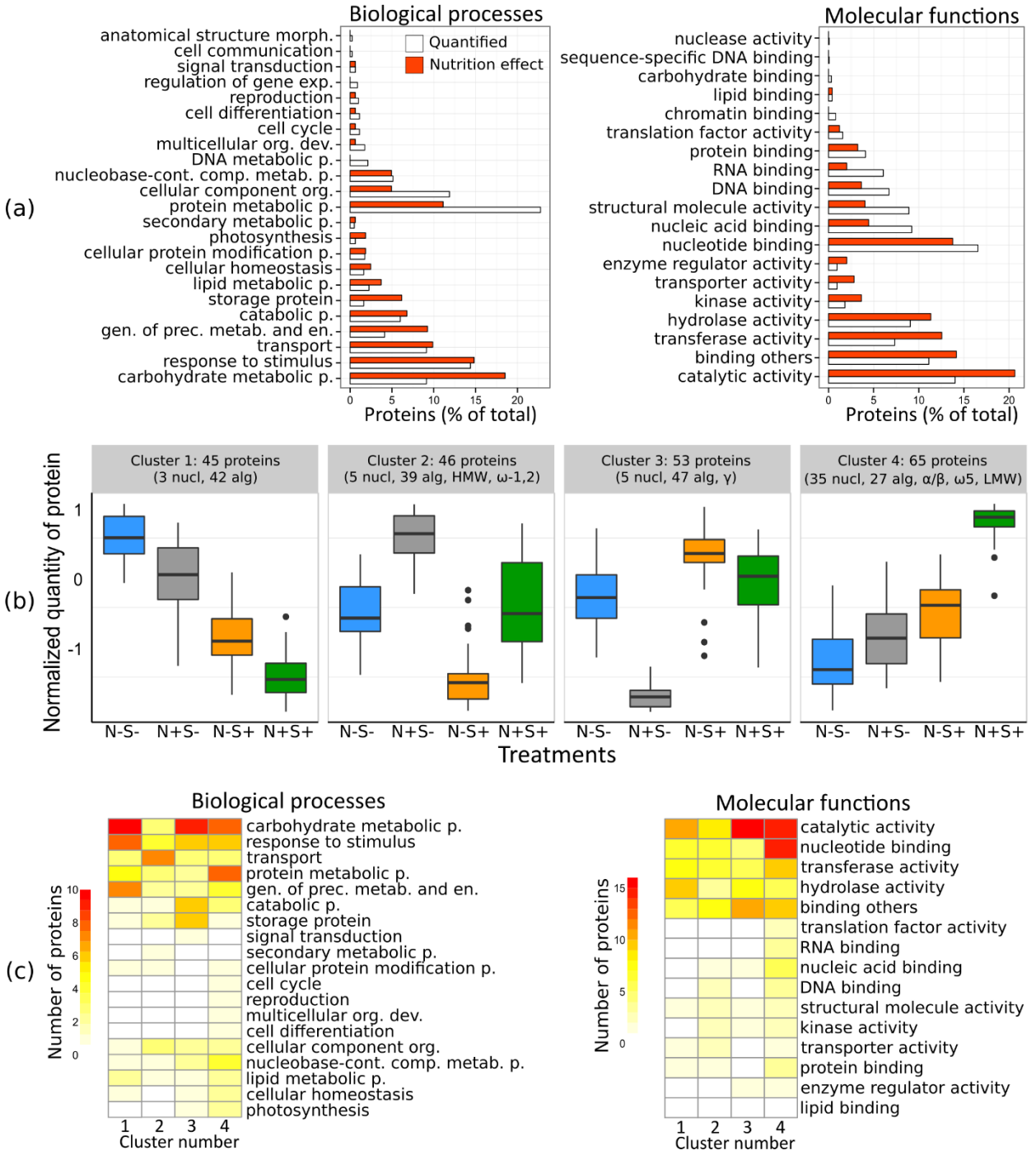


Figure 4. Functions and response to nutrition profiles of proteins showing significant nutrition effect. (a) Percentage of proteins in each biological process and molecular function classes for the 978 quantified proteins (open bars) and for the 209 proteins (155 albumin-globulin proteins, 48 nuclear proteins, and 6 grain storage proteins) which differed in abundance between the treatments (filled bars). (b) Boxplots of the normalized quantity of proteins in the four clusters for the four treatments. Boxes show the 25th to 75th percentile range, horizontal lines in boxes show medians, error bars outside boxes show the 10th to 90th percentile range, dots are outliers. The number of proteins in each cluster is indicated at the top of each panel. alg, albumin-globulin, nucl, nuclear protein. (c) Heatmaps of the number of proteins in different biological process and molecular function classes for the four clusters shown in (b).

Only 69 (7%) of the 978 quantified proteins were present in both fractions, showing again the very low redundancy between the albumin-globulin and nuclear protein analyses (Figure 3c). Of these 69 proteins, 20 (29%) were ribosomal proteins, classified as functioning in protein metabolic processes and structural molecule activity (Figure 3d, e). Other proteins are involved in response to stimulus (11%) or transport processes (8%), with for example two importins (Figure 3d), and many possess nucleotide binding (12%) or RNA binding (9%) activities (Figure 3e).

Various functions and response patterns in proteins impacted by N and S supply

The abundance of 209 proteins (i.e. 21% of the quantified proteins) was significantly affected by the nutrition: the six GSPs, 155 albumin-globulin proteins and 48 nuclear proteins; Table S5). A difference in GO term enrichment was observed between the 978 quantified proteins and the 209 nutrition-responsive proteins. Proteins acting in carbohydrate metabolic processes, generation of precursor metabolites and energy, or storage functions were more represented among the proteins whose abundance was modified by N and/or S supply (Figure 4a). Conversely, protein metabolic process (e.g. many ribosomal proteins) and cellular component organization process (e.g. many histones), which made up a high proportion of the quantified proteome (22.7% and 11.8%, respectively) were less well represented among the proteins whose abundance was modified by N and/or S supply (11.1% and 4.9%, respectively). Albumin-globulin proteins were proportionally more affected by N and S supply (30.2%) than nuclear proteins (10.3%), with enrichment of GO terms for catalytic, transferase, hydrolase or enzyme regulator activities (Figure 4a). Consistent with this result, DNA and RNA binding activities were underrepresented among the proteins whose abundance was nutrition-responsive.

To analyze how the abundance of the 209 proteins responded to N and/or S supply, clustering analysis was performed and 4 clusters were established (Figure 4b). Clusters 1 and 4 grouped proteins with opposite response profiles. Cluster 1 included 45 proteins that were much less abundant in response to N+S+, none of which were GSPs. On the contrary, cluster 4, the largest cluster with 65 proteins, grouped proteins which became more abundant in response to N+S+ and included α/β-gliadin, ω5-gliadin, LMW-GS and 35 (73%) nuclear proteins. Cluster 2 grouped 46 proteins that were highly abundant in the N+S- treated grain and included the S-poor HMW-GS and ω1,2-gliadin. Cluster 3 grouped 53 proteins that were more abundant in response to N-S+ and included the S-rich γ-gliadin. Clusters 2 and 3 indicate a clear difference between N and S imbalances on the grain proteome, as the 99 proteins grouped in these two clusters behaved similarly in response to N-S- and N+S+.

Generally, proteins with similar molecular functions or acting in the same biological process were found in several clusters. However, some processes/functions were preferentially detected in one, two or three clusters (Figure 4c). For example, proteins acting in carbohydrate metabolic processes or in response to stimulus were poorly represented in cluster 2, few of them being increased by high N supply, unlike proteins involved in transport processes. Proteins acting in the generation of precursor metabolites and energy, also classified as acting in carbohydrate metabolic processes, were mainly members of cluster 1 and thus decreased with high N and S supply. Nucleotide binding, RNA binding and DNA binding proteins were mostly found in cluster 4, and were nuclear proteins that mostly increased by high N and S supply.

Network analysis reveals central actors in the grain proteome response to nutrition

To provide an integrative view of the grain response to N and S supply and highlight proteins that co-accumulate during grain filling, a directed multilevel network was constructed using proteomic data for the 209 proteins whose abundance was modified by N and/or S supply from 300 to 500°Cd after anthesis. We used a data mining method based on the discovery of association rules between attributes (Agier *et al.*, 2007; Vincent *et al.*, 2015). Associations between proteins and comparisons between treatments were built according to individual protein amounts in the different samples. The resulting network comprised 206 significant linkages (edges) and 111 proteins (nodes), i.e. about 50% of the 209 input proteins, and included the six GSPs, 25 nuclear proteins, and 80 albumin-globulin proteins (Figure 5a; Table S5). Networked proteins showed significant and high differences in abundance between at least one pair of treatments. The comparison between N+S- and N-S+ treatments created the highest number of edges (61, 29.6%) suggesting that the effect of these treatments on the grain proteome differed the most.

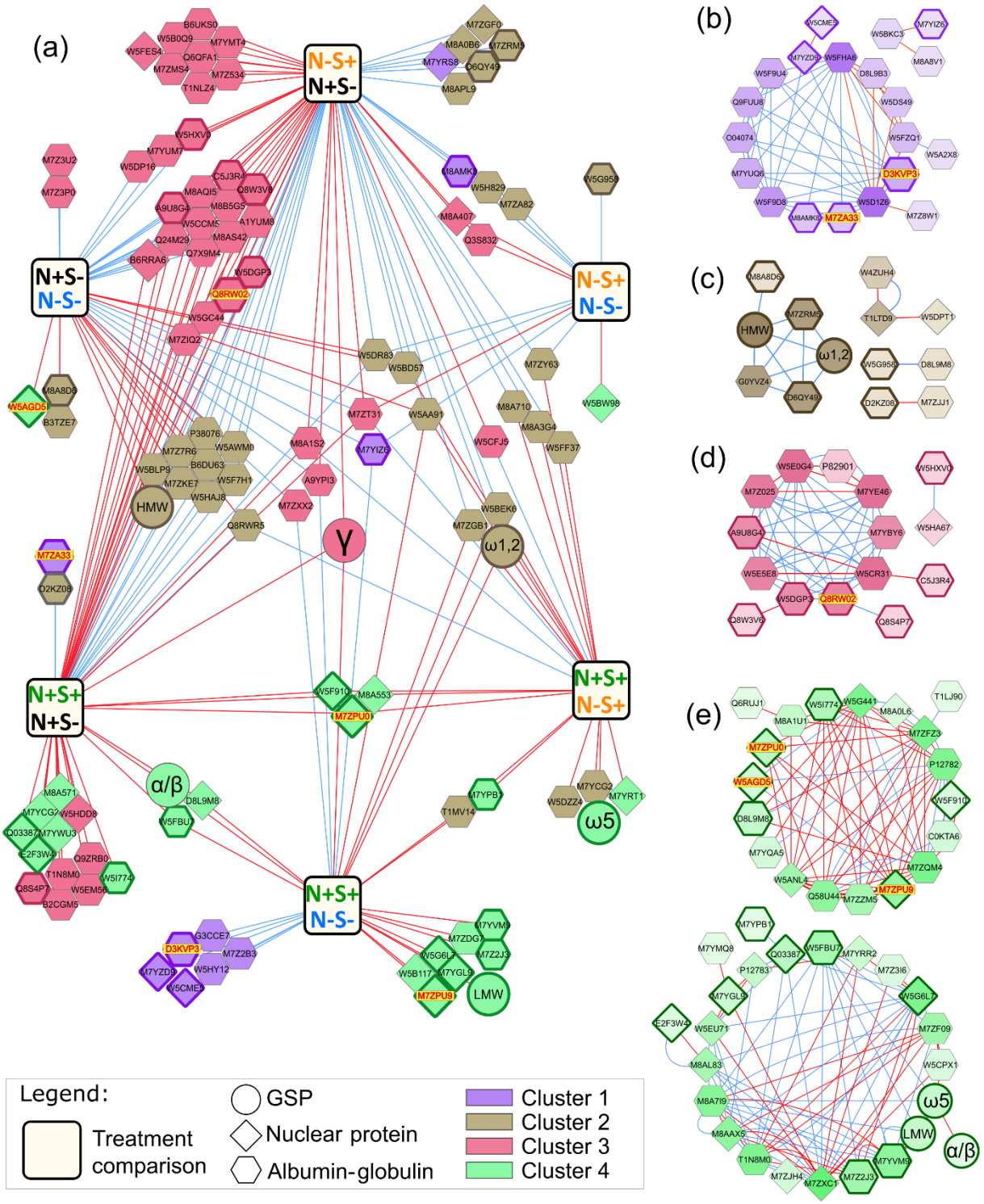


Figure 5. Network analysis of einkorn grain subproteome response to N and S supply. (a) Co-abundance network for proteins which differed in abundance between treatments and linkages with N and S treatments defined as central attributes. (b) to (e) Co-abundance networks for proteins in cluster 1 (b), cluster 2 (c), cluster 3 (d) and cluster 4 (e) as shown in Figure 4. Node shapes reflect attribute categories: squares, comparison between two treatments; circles, grain storage proteins (GSPs); hexagons, albumin-globulin proteins; and diamonds, nuclear protein. Node colours indicate which cluster the protein belongs to: purple, cluster 1; brown, cluster 2; red, cluster 3; green, cluster 4. In (b) to (e) the intensity of the node colour reflects the number of edges (connectivity). Edge colour indicates the biological significance between source and target: in (a) red when the protein was more abundant in the first treatment used in the comparison and blue when the protein was less abundant in the first treatment; in (b, c, d, e), red shows the abundance of proteins was high for the two nodes and blue when the abundance of proteins was low for the two nodes. Nodes with thick borders were found both in (a) and a corresponding cluster network in (b), (c), (d), or (e). Albumin-globulin and nuclear proteins are identified by their Uniprot accession reference (for details, see Table S6). Accessions written in red correspond to proteins highlighted in Figure 6.

The comparison between N+S+ and N+S- was involved in 49 rules, highlighting the importance of N × S interactions on the grain proteome compared to N alone. Only 27 proteins differed in abundance between N+S+ and N-S-. This result demonstrates the importance of the balance between N and S supply. N+S-/N-S-, N-S+/N-S- and N+S+/N-S+ pairwise comparisons linked 35, 12 and 22 proteins, respectively.

In this network, many groups of proteins were formed. Grouped proteins had the same response to N and S supply and were mostly part of the same cluster, an indicator of the robustness of the network analysis. For example, the quantity of LMW-GS was higher for N+S+ compared with N-S- and four nuclear proteins and three albumin-globulin proteins were also grouped with LMW-GS: an RNA helicase 27, an alanine aminotransferase, a betaine dehydrogenase, a guanine nucleotide-binding 3 homolog, a monothiol glutaredoxin-S11, a putative NADP-dependent oxidoreductase P1 and a chitinase 1. Some proteins were linked because they had very different quantities in three or four treatment comparisons and were therefore strongly affected by N and S supply. For example, HMW-GS was abundant with N+S- treatment compared with the three other treatments. Several albumin-globulin with similar responses to N and S supply were thus grouped with HMW-GS: a zeta-carotene desaturase, a cysteine synthase, a glutathionyl-hydroquinone reductase, an isoflavone reductase, a methylthioribose kinase and a beta-D-glucan exohydrolase. Other proteins were highly connected in this network, for example, γ- and ω1,2-gliadin and several albumin-globulin proteins, like an importin subunit alpha, which was connected with four edges. This importin subunit was more abundant for N+S- and N+S+ than with N-S- and N-S+, showing that this protein is increased by high N availability independently of S supply.

Co-accumulation of proteins was also investigated in other networks, built within cluster (Figure 5b-e). In cluster 1 co-accumulation network, there were 20 (44%) proteins (nodes) including four involved in carbohydrate metabolic process, two acting in response to stress, two ribosomal proteins, a serpin-ZX and a purple acid phosphatase (Figure 5b). In cluster 2 co-accumulation network, there were 13 (28%) proteins including HMW-GS and ω1,2-gliadins (Figure 5c). According to the network, the abundance of these GSPs was connected to that of two other storage proteins and a programmed cell death protein 4. The network connected an importin subunit beta-1-like and a tyrosine-protein phosphatase with two edges (abundant in similar conditions and not abundant in other similar conditions), showing that these two proteins strongly co-accumulate. The co-accumulation network from cluster 3 contains 15 (28%) nodes, including two vicilin-like antimicrobial peptides 2-2, a cationic peroxidase SPC4-like, an ankyrin repeat domain-containing protein and a glutathione transferase, which were related to the accumulation of other proteins (Figure 5d). The co-accumulation network with cluster 4 proteins is the largest, meaning that proteins that increased with N+S+ co-accumulated in tight synchrony (Figure 5e). This network was composed of two modules. One module contained 19 nodes (i.e. 29% of the proteins of cluster 4) and 70 edges, the other one contained 24 nodes (37%) and 85 edges. Some proteins in the first module were highly connected to others, including a RNA helicase 27, a L-ascorbate peroxidase, a phosphoglycerate kinase, a RNA polymerase and a nucleolar complex protein 3. The second module comprised LMW-GS, ω5- and α/β-gliadin and several other co-accumulating proteins: a betaine aldehyde dehydrogenase, a probable aldo-keto reductase 1, a phosphorylase and a multiprotein-bridging factor 1a.

In these two network approaches several proteins were revealed as being strongly affected by N and/or S supply and co-accumulating with other proteins (nodes with thick border line in Figure 5; Table S5). These proteins thus represent central actors of the proteome response to nutrition.

Focus on central actors of the einkorn grain proteome response to N and S supply

Among proteins revealed in the network analysis, a DEAD-box ATP-dependent RNA helicase 27 (M7ZPU9) was found in cluster 4 (Figure 6a). This protein is known to be involved in many processes including RNA transcription, pre-mRNA splicing, mRNA export, ribosome biogenesis and translation initiation (Linder, 2006). This RNA helicase responded like LMW-GS to N and S supply, and was connected in the co-accumulation network to 11 other proteins, including a trihelix transcription factor (Figure 6b) and a peter-pan like protein (Figure 6c). The trihelix transcription factor contains a SANT/Myb domain, like those found in chromatin-remodeling complexes (Boyer *et al.*, 2004), and its quantity was linked in network with that of four proteins (Figure 5e).

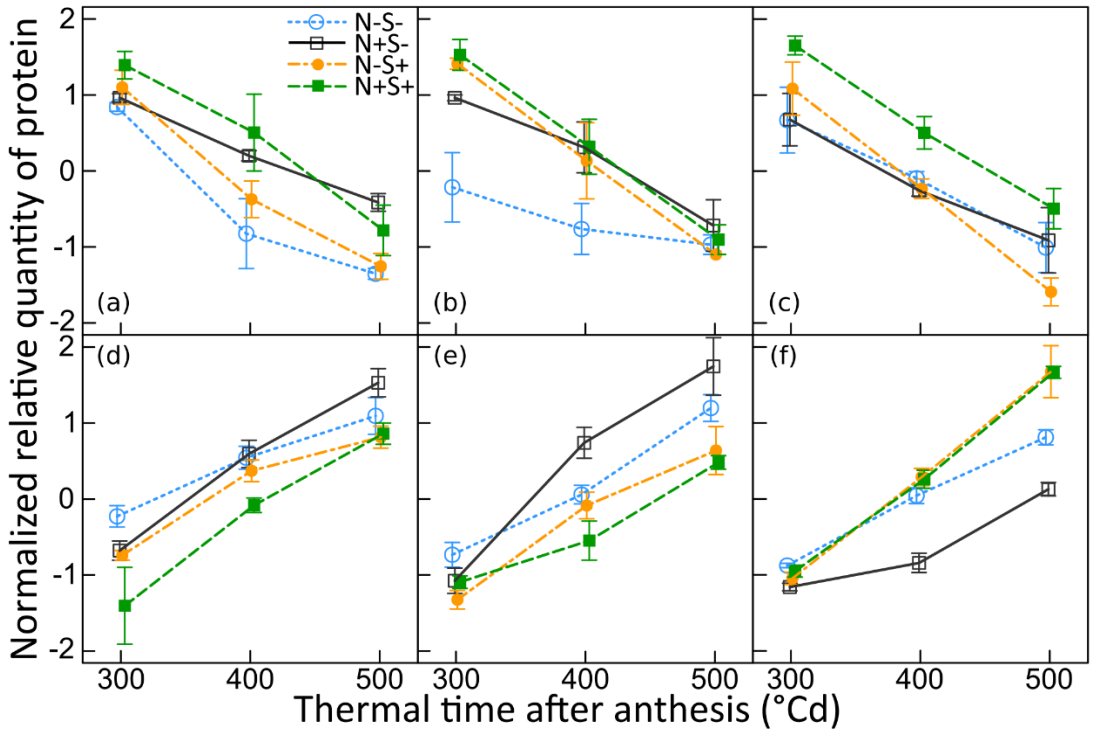


Figure 6. Effects of N and S supply on the quantity of selected nuclear and albumin-globulin proteins during einkorn grain filling. (a) DEAD-box ATP-dependent RNA helicase 27, (b) trihelix transcription factor, (c) peter pan-like protein, (d) purple acid phosphatase, (e) serpin-ZX, (f) glutathione transferase. Data are scaled and are means ± 1 s.e. for $n = 3$ independent replicates.

The peter-pan like protein possesses a Brix domain, found in proteins involved in ribosomal RNA processing (Weis *et al.*, 2015) and was networked with six proteins (Figure 5e). These three central actors co-accumulated and all increased with N and S supply. Two other central actors were part of cluster 1 and thus decreased with N and S supply: a purple acid phosphatase (Figure 6d) acting in protein post-translational modification and a serpin-ZX (Figure 6e) that negatively regulates endopeptidase activity. The amounts of these two proteins were associated with those of seven and four other proteins, respectively, including a probable sarcosine oxidase. Like γ-gliadin, the abundance of two glutathione transferases increased markedly with high S, one of which was central in the network being connected to nine other proteins (Figures 5d and 6f).

DISCUSSION

In wheat, N and S supply highly influences GSP accumulation, and is thus expected to impact the viscoelastic properties of gluten and flour end-use value. There is currently a concerted effort to reduce the use of N-based fertilizers and soil S availability is decreasing in many wheat growing regions. It is in this context that we describe grain GSP, nuclear and albumin-globulin subproteome responses to N and S supply in a diploid species closely related to bread wheat and durum wheat (*T. turgidum* L. ssp. *durum* (Desf.) Husn.). The 203 albumin-globulin and nuclear proteins whose abundance was modified by N and/or S supply have diverse functions. Clustering and network analyses allowed us to group proteins that had similar abundance profiles in response to N and/or S supply during grain filling. We found that S supply highly impacted GSP accumulation while N and S supply had different effects on the grain proteome. Several proteins central to the proteome response and potentially involved in the regulation of GSP synthesis in response to nutrition were highlighted. This information is summarized in a scheme of grain metabolism illustrating which molecular pathways are affected by N and S supply during grain filling (Figure 7).

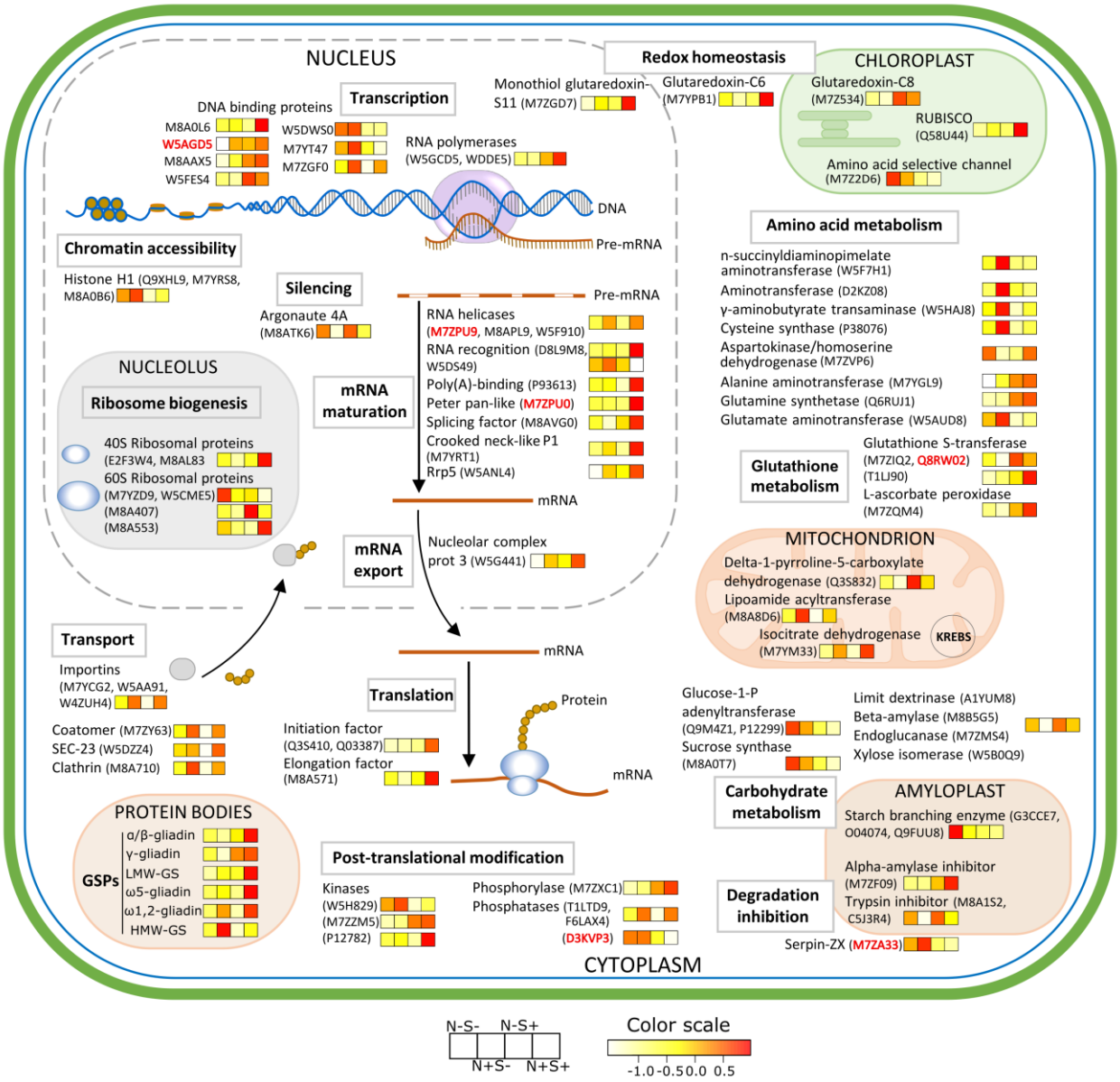


Figure 7. Schematic overview of einkorn grain subproteome responses to N and S supply. Proteins represented are some of the 209 proteins showing significant nutrition effects during grain filling. Relative protein quantities are scaled and means for n = 3 independent replicates are represented by a color scale on a 1 × 4 grid, each square representing one treatment. Functional classification was based on gene ontology annotation and subcellular localization was determined by prediction tools. Uniprot accessions corresponding to protein descriptions are given in parentheses. Accessions in red correspond to proteins highlighted in Figure 6.

The strong influence of S on the quantity and composition of grain storage protein and other grain subproteomes

S deficiency in soil leads to decreases in the amount of GSP in wheat grain (Shewry *et al.*, 2001). We obtained similar results with more total GSP per grain at maturity in the two S-containing treatments than in the low S treatments. This was due to an early increase in the rate of accumulation and longer duration of accumulation of gliadin, especially the S-rich classes α/β- and γ-gliadin. This is in accordance with previous reports that the highest amount of grain gliadin was obtained with moderate N fertilization and high S supply (Zörb *et al.*, 2010).

In the present study, the abundance of three glutathione transferases was higher in response to N-S+ and N+S+ treatments than to N-S-. One of them was also identified as a central actor and had similar nutrition response profiles to those of γ-gliadin. This result suggests that glutathione may increase with S supply, as occurs in wheat (Dai *et*

al., 2015). Glutathione is one of the major pools of non-protein S in wheat grain (Rhazi *et al.*, 2003; Steinfurth *et al.*, 2012). In S-deficient conditions, lesser amounts of total cysteine and glutathione were found in wheat flour, which affected GSP composition and hence technological flour properties (Reinbold *et al.*, 2008). Glutathione is involved in the glutathione-ascorbate cycle controlling the concentration of active oxygen in the cell (Noctor and Foyer, 1998). In this mechanism, ascorbate peroxidase is involved in the detoxification of H_2O_2 (Caverzan *et al.*, 2012). One ascorbate peroxidase was also more abundant in einkorn grain treated with S. It is therefore possible that glutathione metabolism increases when S is freely available.

In network analysis, N+S- and N-S+ had the most contrasting effects on the grain proteome, since three GSPs (HMW-GS, ω1,2- and γ-gliadin) and many nuclear and albumin-globulin proteins differentially accumulated between these two treatments. N+S+ and N+S- also had very different effects, again demonstrating the strong effect of S on the proteome. Interestingly, even though many proteins were increased with N+S+, the network analysis of the N+S+/N-S- comparison connected fewer proteins than the N+S-/N-S+ comparison (27 and 61 proteins, respectively). Taken together, these results suggest that the grain proteome responded more to the N/S balance than to the availability of either N or S individually. These results are in good agreement with previous studies showing the importance of N-to-S ratio in determining the rheological properties of wheat dough and GSP composition (Zhao *et al.*, 1999; Zörb *et al.*, 2010; Dai *et al.*, 2015).

Grain metabolism was influenced by N and S supply

Carbohydrate metabolism was the biological process most represented in proteins with significant nutrition effects. Such proteins were found in all four of the clusters established, but a few of them were in cluster 2, that is they didn't increase with N only. Conversely, proteins involved in amino acid metabolism were highly abundant with high N supply. This is consistent with previous results where an increased concentration of free amino acids in wheat flour was measured in high N-to-S ratio conditions (Granvogl *et al.*, 2007). We noted that the abundance of a cysteine synthase increased considerably with high N supply, which tallies with a previous observation in rice (*Oryza sativa* L.) shoots and roots where high levels of cysteine synthase gene transcripts were observed in S starvation and non-limiting N conditions (Nakamura *et al.*, 1999). The latter authors concluded that activation of this gene regulated the balance between S-containing amino acids (methionine and cysteine) and other amino acids. These results broadly indicate that amino acid metabolism is activated at the expense of sugar metabolism in response to N nutrition.

Intracellular transport was likely influenced by nutrition. Indeed, transport was one of the few biological processes strongly enriched in protein cluster 2. Proteins participating in transport were more abundant with high N supply and less abundant with high S supply. Among transporter proteins significantly affected by nutrition were three importins, two of which were present in both the nuclear and albumin-globulin protein extracts. Importins are involved in the transport of NLS-containing proteins from the cytoplasm to the nucleus through nuclear pore complexes localized in the nuclear envelope (Bednenko *et al.*, 2003). In plants, communication between these two cell compartments is important for growth, development and responses to environmental stimuli (Merkle, 2011; Tamura and Hara-Nishimura, 2013). Nucleocytoplasmic transport may be a regulatory mechanism activated by grain N/S status to modify transcription by activating import of transcriptional regulators.

Nutrition impact on DNA binding proteins suggests transcriptional reprogramming

In wheat, GSP synthesis is regulated at the transcriptional level (Dai *et al.*, 2015). This regulation, first described in barley and conserved in other cereals, is governed by a network involving cis-regulatory elements located in promoters of GSP genes and their interacting transcription factors (Rubio-Somoza *et al.*, 2006; Verdier and Thompson, 2008). The eight transcription factors known to act in this regulation were not quantified in our experiment, which didn't allow us to say if their abundance was disturbed by N and S supply. Among the 48 nuclear proteins showing significant nutrition effects, 10 (21%) have a DNA binding function. Proportionally then nutrition had a strong effect on proteins with DNA binding domains as this molecular function represented only 10% of the quantified nuclear proteins. Among them, two DNA-directed RNA polymerases, responsible for RNA synthesis, were increased by N and S supply. A similar response was observed for a multiprotein-bridging factor 1a. In *Arabidopsis thaliana*, this

type of factor is known to be a transcriptional co-activator (Tsuda *et al.*, 2004). These results suggest that transcription is activated under conditions of high N and S availabilities. As might be expected if more transcripts are being produced, several proteins involved in mRNA splicing also accumulated when N and/or S were provided.

In the DNA-binding domain class were three histone H1 proteins. Histones are responsible for DNA condensation, organization and regulation in the nucleus. In eukaryotes, assembly of two copies of each of the core histone, H2A, H2B, H3 and H4, results in an octamer protein that binds to superhelical DNA to form the basic structure of DNA compaction called the nucleosome. Histone H1 molecules are linkers acting to maintain chromatin structure (Harshman *et al.*, 2013). Three variants of histone H1 were part of clusters 1 and 2, i.e. they were less abundant when S was supplied. In *A. thaliana*, histone H1 mutations resulted in modifications of DNA methylation patterns, with more methylation found in some promoter sequences in endosperm (Wierzbicki and Jerzmanowski, 2005; Rea *et al.*, 2012). In plants as in animals, DNA methylation affects the binding of specific proteins to DNA and therefore the formation of the transcription machinery (Finnegan *et al.*, 1998; Vanyushin and Ashapkin, 2011). Interestingly in our study, there was one protein with a methyl-CpG DNA binding domain (MBD) in cluster 3, i.e. it increased with high S supply. MBD proteins potentially act in transcriptional repression by recruiting chromatin remodeling factors, histone methyltransferases and histone deacetylases, leading to chromatin compaction (Grafi *et al.*, 2007). These results suggest that an epigenetic response to S supply occurs in einkorn grain. A trihelix transcription factor that increased in the high N and high S condition was predicted to be a central actor in the grain proteome response. This uncharacterized protein contains a SANT/Myb domain, characteristic of some nuclear receptor co-repressors and in many chromatin-remodeling complexes (Boyer *et al.*, 2004). In barley, it has been reported that methylation of the promoters of storage protein genes could repress storage protein synthesis and control grain development (Sørensen *et al.*, 1996; Radchuk *et al.*, 2005). Thus the methylation state of GSP promoters could be a regulatory mechanism, modifying the DNA-binding capacity of transcription factors, as in maize Opaque 2 transcription (Rossi *et al.*, 1997; Sturaro and Viotti, 2001; Locatelli *et al.*, 2009). Taken together, these different elements indicate that N and S alter transcription dynamics, with potential effects on chromatin compaction and accessibility governed by epigenetic mechanisms.

In summary, major changes occur in the proteome of developing einkorn wheat grain when the nutritional status is modified. Post-flowering N and S nutrition clearly influenced the N-to-S ratio in the grain, leading to significant changes in the GSP composition in mature grain, by modification of the rate and duration of GSP accumulation, especially of HMW-GS, α/β- and γ-gliadin. Changes in N-to-S ratios were also related to changes in the nuclear and albumin-globulin proteomes. The balance of cell functions was probably disturbed by altering N and S supply, as albumin-globulin proteins involved in carbohydrate and amino acid metabolism and transport were differentially affected. Central to these major changes, several DNA-binding nuclear proteins could have roles in the early grain response. Several proteins highlighted in the present study will be targeted to investigate how grain composition is controlled.

EXPERIMENTAL PROCEDURES

Plant material and growth conditions

Triticum monococcum ssp. *monococcum* (accession ERGE 35821) seeds were germinated (5 February) at room temperature on filter paper moistened with demineralized water in Petri dishes. When the radicles were 0.5 cm to 1 cm long, seedlings were transplanted to 50 mL PVC columns (7.5 cm internal diameter (i.d.) × 50 cm deep; 2 seedlings per column) filled with a 2:1 (v/v) mixture of washed perlite and river sand. The columns were arranged in a greenhouse to form a homogenous stand with a population density of 512 plants m^{-2}. The experimental design was a randomized complete block design with four blocks and four treatments. The high plant density inhibited the development of axillary tillers which favored the synchronous development of the plants within and between containers. In the greenhouse, air temperature at the top of the plant stand was maintained at 20°C/15°C (16 h light/8 h dark) and air relative humidity at 55%/75%. During light periods plants received a mean total daily photosynthetic photon flux (PPF) of 152 and 161 µmol m^{-2} d^{-1} between transplantation and anthesis and between anthesis and grain ripeness maturity, respectively. Air temperature and relative humidity were measured in miniature forced-draft flues placed in each block in the center of the plant stands at ear height. Air temperature, relative humidity, and PPF were measured every min and 15-min

averages were recorded with a data logger. Thermal time was calculated as cumulative degree-days (0°C base temperature) using average daily air temperature.

Until anthesis, each PVC column received 167 mL d^{-1} of a modified Hoagland's nutrient solution (Castle and Randall, 1987) containing 3 mM N and 0.1 mM S prepared with demineralized water (1 mM KH_2PO_4, 1 mM KNO_3, 0.5 mM $Ca(NO_3)_2$, 0.5 mM NH_4NO_3, 0.1 mM $MgSO_4$, 1.9 mM $MgCl_2$, 3.5 mM $CaCl_2$, 4 mM KCl, 10 µM H_3BO_3, 0.7 µM $ZnCl_2$, 0.4 µM $CuCl_2$, 4.5 µM $MnCl_2$, 0.22 µM MoO_3 and 50 µM EDFS-Fe). At anthesis when N and S demand is low, the nutrient solution was replaced with demineralized water to avoid excess build-up of N and S compounds in plants or potting substrate. Then from 200°Cd to 700°Cd after anthesis four combinations of N and S were supplied: N-S-, nutrient solution with no N or S; N+S-, 6 mM N with no S; N-S+, low N (0.5 mM) and high S (2 mM); or N+S+, high N (6 mM) and high S (2 mM). Nitrogen was not totally excluded from the N-S+ nutrient solution as previous experiments have shown that S uptake and/or metabolism are inhibited when N is absent from the nutrient solution (P. Martre, unpublished data). The nutrient solutions were modified as described in Dai *et al.* (2015).

Grain sampling and processing

Main-stem ears were tagged when the anthers of the central florets appeared. Grains were collected from the central portion of ears every 100°Cd (approximately every 5 d) from 300°Cd after anthesis to ripeness. For each treatment and sampling date, four to ten main-stem ears were sampled per replicate, depending on the grain developmental stage. Except for those used to determine grain dry mass, grains were frozen in liquid N_2 just after harvesting then stored at -80°C. Four biological replicates of whole grain were analyzed unless indicated otherwise.

Determination of total grain S and N concentration

Aliquots of 5 mg of wholemeal flour were weighed in tin capsules and the total N and S concentrations were determined with the Dumas combustion method using a FlashEA 1112 NC Analyzer (Thermo Electron) following the manufacturer's recommendations. Grain protein content was calculated by multiplying grain N concentration by 5.62 (Mossé *et al.*, 1985).

Storage protein extraction and quantification

Gliadin and glutenin proteins were sequentially extracted from 100 mg of wholemeal flour milled from each replicate of grain sampled between 300°Cd and 1000°Cd after anthesis, as described by Plessis *et al.* (2013). Gliadin classes and glutenin subunits were separated and quantified by reverse phase high performance liquid chromatography (RP-HPLC) using an Agilent 1290 Infinity LC system (Agilent Technologies, http://www.agilent.com) as described by Dai *et al.* (2015). Briefly, gliadin and glutenin extracts were filtered through regenerated cellulose syringe filters (0.45 µm pore diameter, UptiDisc, Interchim, http://www.interchim.com), then 4 µL (for gliadins) or 2 µL (for glutenins) of protein extract were injected into a C8 reversed-phase ZORBAX 300StableBound column (2.1 × 100 mm, 3.5 µm, 300 Å; Agilent Technologies) maintained at 50°C. Proteins were separated at a flow rate of 1 mL min^{-1} by using linear solvent gradients from 24% to 50% acetonitrile containing 0.1% (v/v) trifluoroacetic acid over 13 min for gliadins, and from 23% to 42% over 25 min for glutenins. Proteins were detected by UV absorbance at 214 nm. Chromatograms were processed with ChemStation 10.1 software (Agilent Technologies) and the HPLC peaks corresponding to each of the four gliadin classes and the two glutenin subunits were identified following the observations of Wieser *et al.* (1998; Figure S2).

Nuclear and albumin-globulin protein extraction, digestion and desalting

Nuclei were extracted and purified from 2 g of whole grains ground in extraction buffer (20 mM Hepes-KOH pH 7, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.5 mM PMSF, 0.1% (v/v) phosphatase inhibitor cocktail (Sigma-Aldrich)) with a Polytron homogenizer (Kinematica POLYTRON® PT 10) according to Bancel *et al.* (2015). Nuclear proteins were then prepared using TRI Reagent® (Sigma-Aldrich) according to the manufacturer's instructions. Albumins and globulins were extracted from whole grains according to Marion *et al.* (1994) with the following modifications. Proteins were extracted for 2 h at 4°C in 10 mM sodium phosphate, 10 mM NaCl, pH 7.8. After centrifugation at 8000 × *g* for 20 min at 4°C, proteins in the supernatant were precipitated with ice-cold acetone for 2 h at -20°C. After centrifugation at 10 000 × *g* for 5 min at 4°C, the resulting pellets were washed three times in ice-cold acetone then dried at room temperature.

Nuclear and albumin-globulin protein pellets were solubilized in 30 µL of a buffer containing 0.1% (v/v) ZALS I, 6 M urea, 2 M thiourea, 10 mM DTT, 30 mM Tris-HCl pH 8.8, and 5 mM NH_4HCO_3. Protein concentration was measured using the 2-D Quant Kit (GE Healthcare) with bovine serum albumin as standard. For each sample, 40 µg of protein was alkylated with 58 mM iodoacetamide for 50 min in the dark at room temperature, then diluted 10-fold with 50 mM NH_4HCO_3 and digested in solution overnight at 37°C with 800 ng of trypsin in 50 mM NH_4HCO_3. Digestion was stopped by adding 10% TFA.

The protein samples were desalted using a Strata-XL SPE column (Phenomenex) regenerated with 500 µL of 100% CH_3CN and equilibrated three times with 500 µL of buffer A (3% CH_3CN, 0.06% acetic acid). Digested protein (15 µg) mixed with buffer A was loaded onto the column. After three washes with 500 µL of buffer B (40% CH_3CN, 0.06% acetic acid), peptides were eluted twice with 300 µL of buffer B, then dried and stored at -20°C.

LC-MS/MS analysis of nuclear and albumin-globulin proteins

Peptide samples were solubilized in a buffer of 3% CH_3CN, 0.05% TFA and 0.05% formic acid. Liquid chromatography was performed on a NanoLC Ultra system (Eksigent). Samples (1 µg) were loaded at 7.5 µL min^{-1} on a C18 precolumn (5 µm, 100 µm i.d. × 2 cm length; NanoSeparations) connected to a separating BIOSPHERE C18 column (3 µm, 75 µm i.d. × 150 mm length or 300 mm length for nuclear and albumin-globulin proteins, respectively; NanoSeparations). Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in CH_3CN. Peptide separation was achieved using a linear gradient from 5% to 35% solvent B for 28 min (for nuclear proteins) or 60 min (for albumin-globulin) at 300 nL min^{-1}. Including the regeneration and the equilibration steps, a single run took 45 min for nuclear proteins and 87 min for albumin-globulin proteins. Eluted peptides were analyzed with a Q-Exactive mass spectrometer (Thermo Electron) using a nanoelectrospray interface. Ionization was performed with a 1.3 kV spray voltage applied to an uncoated capillary probe (10 µm i.d., New Objective). The Xcalibur 2.3 SP1 interface was used to monitor data-dependent acquisition of peptide ions. This included a full MS scan covering a mass-to-charge ratio (m/z) of 400 to 1,400 with a resolution of 70,000 and an MS/MS step (normalized collision energy, 27%; resolution, 17,500). The MS/MS step was reiterated for the eight major ions detected during the full MS scan. Dynamic exclusion was set to 40 s.

Peptide identification and quantification

Nuclear and albumin-globulin proteins were identified by matching peptides against the Uniprot protein database version 2014_07 limited to the *Triticum* genus using X!Tandem version Sledgehammer 2013.09.01.1 (www.thegpm.org/tandem). Enzymatic cleavage parameters were set as trypsin digestion with one possible missed cleavage. Cys carboxyamidomethylation was set as static modification, whereas Met oxidation, N-terminal deamidation, and N-terminal acetylation were set as variable modifications. Precursor mass tolerance was 10 ppm and fragment mass tolerance was 0.02. Identified proteins were filtered and grouped using X!TandemPipeline version 3.3.4 (http://pappso.inra.fr/bioinfo/xtandempipeline/). Peptide and protein e-value cut offs were set to 0.01 and 10^{-4}, respectively, with at least two peptides per protein.

Relative quantification of all identified peptides was done using MassChroQ software version 2.1.3 (Valot *et al.*, 2011) and extracting ion chromatograms (XICs) for all identified peptides and integrating the area of XIC peaks at the corresponding retention time. Non repeated peptides were then removed. For the albumin-globulin fraction, these corresponded to peptides present in less than 90% of the samples. For nuclear proteins, this threshold was set at 30%, due to the high effect of the developmental stage on this subproteome, as previously reported on wheat grain (Bonnot *et al.*, 2015). To quantify proteins, only peptides not shared by other proteins and correlated peptides were used. The quantity of the corresponding protein was calculated by summing the quantities of specific peptides for albumin-globulin, and by averaging the quantities of specific peptides for nuclear proteins.

Functional classification and subcellular localization

Identified nuclear and albumin-globulin proteins were classified according to gene ontology (GO; Ashburner *et al.*, 2000). Proteins that differed in quantity between treatments but did not have any functional annotation or GO information were analyzed with Blast2GO tool version 3.2 (Conesa *et al.*, 2005) in order to assign potential function.

Predicted subcellular localization using MultiLoc2 (Blum *et al.*, 2009), WoLF PSORT (Horton *et al.*, 2007) and LocTree3 (Goldberg *et al.*, 2014) programs were used to validate nuclear protein identification. Results were collected using the integrative tool PSI (Liu *et al.*, 2013) which combines prediction results of different programs. The top three hits were considered for WoLF PSORT and MultiLoc2. Nuclear proteins were used for statistical analysis if they were predicted to be nuclear with at least one tool.

Statistical analysis

Data were analyzed using the statistical software program R v3.2.2 (R Core Team, 2015). Differences in grain dry mass per grain and per ear, total quantity of N and S per grain, and the quantity of GSP per grain at maturity were analyzed using a one-way ANOVA with the factor nutrition (four levels). For the quantity of GSP, albumin-globulin and nuclear proteins in immature grains, a two-way ANOVA was performed, using the factors of nutrition and developmental stage (i.e. the thermal time after anthesis). Significant differences (α = 0.05) were then analyzed by Tukey's honestly significantly different (HSD) post-hoc test.

To determine the rate and duration of accumulation of grain dry mass, total N and S and protein fractions, data were fitted with a 3-parameter logistic function equation (Triboï *et al.*, 2003):

$$Q(t) = \frac{Q\max}{1 + 0.05\exp\left(\frac{-4r(t - t_{95})}{Q\max}\right)} \quad (1)$$

where Q is the quantity of dry mass, N, S, or GSP per grain, t is the thermal time after anthesis in degree-days, $Q\max$ is the final value of Q approached as $t \to \infty$, r is the maximum rate of accumulation defined as the derivative of the point of inflexion, and t_{95} is the duration of accumulation defined as the period from anthesis onwards during which 95% of $Q\max$ is accumulated.

Hierarchical clustering on principal components was computed on proteins with significant nutrition effects using the FactomineR package for R (Husson *et al.*, 2015), with the hcpc() function and the 'ward' method. The number of clusters was determined with the suggested partition, i.e. the one with the higher relative loss of inertia.

Proteomic network inference and analysis

Relationships among GSP, albumin-globulin and nuclear protein dynamics were analyzed using the RulNet platform for network inference (http://rulnet.isima.fr; Vincent *et al.*, 2015). RulNet uses an algorithm describing rule semantics between attributes (Agier *et al.*, 2007) to generate rules among data. Protein data were scaled and semantics were written to find rules between proteins whose accumulation was significantly affected by nutrition and the four N/S treatments according to the relative abundance of proteins in the different samples (semantic 1, Method S1). Two quality measures were used to select the best rules (confidence and lift). Rules with a confidence > 0.6 and a lift > 1.5 were selected. A second semantic was used to find rules between co-accumulated proteins (semantic 2, Method S1). For this second semantic, thresholds for support, confidence and lift were set at 0.3, 0.9 and 1.5, respectively. Validated rules were visualized using Cytoscape software version 3.3.0 (Smoot *et al.*, 2011).

ACKNOWLEDGEMENTS

We thank Richard Blanc (UMR GDEC, INRA, Blaise Pascal University) for help in the greenhouse and Mireille Dardevet, Annie Faye, Marielle Merlino, Sibille Perrochon and Isabelle Nadaud for help collecting grains (UMR GDEC). This work was supported by a Ph.D. grant from the French Ministry for Higher Education and Research to TB and funding from the French Government managed by the National Research Agency (ANR) in the framework of Investments for the Future (ANR-10-BTBR-03), France AgriMer and the French Fund Supporting Plant Breeding (FSOV).

SUPPORTING INFORMATION

Figure S1. Effects of N and S supply on grain storage protein composition at maturity.
Figure S2. HPLC chromatograms of gliadin (a) and glutenin (b) in mature einkorn grain.
Table S1. Quantity of dry mass, N, S and GSP per grain at maturity.
Table S2. Estimated maximum rate and duration of accumulation of grain dry mass, total N, total S and GSP.
Table S3. Dry mass, N mass and concentration, S mass and concentration, N-to-S ratio, and protein composition of developing einkorn grain.

Table S4. Quantification of nuclear proteins and albumin-globulin in einkorn wheat grain.
Table S5. Summary of the 209 proteins significantly impacted by N and S supply.
Method S1. Semantics used to infer regulatory network from proteomics data.

REFERENCES

Agier, M., Petit, J.M. and Suzuki, E. (2007) Unifying framework for rule semantics: Application to gene expression data. *Fundam. Informaticae*, 543–559.

Ashburner, M., Ball, C.A., Blake, J.A., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

Bancel, E., Bonnot, T., Davanture, M., Branlard, G., Zivy, M. and Martre, P. (2015) Proteomic approach to identify nuclear proteins in wheat grain. *J. Proteome Res.*, **14**, 4432–4439.

Bednenko, J., Cingolani, G. and Gerace, L. (2003) Nucleocytoplasmic transport: navigating the channel. *Traffic*, **4**, 127–135.

Blum, T., Briesemeister, S. and Kohlbacher, O. (2009) MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics*, **10**, 274.

Bonnot, T., Bancel, E., Chambon, C., Boudet, J., Branlard, G. and Martre, P. (2015) Changes in the nuclear proteome of developing wheat (*Triticum aestivum* L.) grain. *Front. Plant Sci.*, **6**.

Boyer, L.A., Latek, R.R. and Peterson, C.L. (2004) Opinion: The SANT domain: a unique histone-tail-binding module? *Nat. Rev. Mol. Cell Biol.*, **5**, 158–163.

Branlard, G., Dardevet, M., Saccomano, R., Lagoutte, F. and Gourdon, J. (2001) Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica*, **119**, 59–67.

Castle, S. and Randall, P. (1987) Effects of sulfur deficiency on the synthesis and accumulation of proteins in the developing wheat seed. *Aust. J. Plant Physiol.*, **14**, 503–516.

Caverzan, A., Passaia, G., Rosa, S.B., Ribeiro, C.W., Lazzarotto, F. and Margis-Pinheiro, M. (2012) Plant responses to stresses: role of ascorbate peroxidase in the antioxidant protection. *Genet. Mol. Biol.*, **35**, 1011–1019.

Chope, G.A., Wan, Y., Penson, S.P., Bhandari, D.G., Powers, S.J., Shewry, P.R. and Hawkesford, M.J. (2014) Effects of genotype, season, and nitrogen nutrition on gene expression and protein accumulation in wheat grain. *J. Agric. Food Chem.*, **62**, 4399–4407.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

Corbellini, M., Empilli, S., Vaccino, P., Brandolini, A., Borghi, B., Heun, M. and Salamini, F. (1999) Einkorn characterization for bread and cookie production in relation to protein subunit composition. *Cereal Chem.*, **76**, 727–733.

Dai, Z., Plessis, A., Vincent, J., et al. (2015) Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. *Plant J.*, **83**, 326–343.

Das, P., Nutan, K.K., Singla-Pareek, S.L. and Pareek, A. (2015) Understanding salinity responses and adopting "omics-based" approaches to generate salinity tolerant cultivars of rice. *Front. Plant Sci.*, **6**, 712.

Eriksen, J. (2009) Chapter 2 soil sulfur cycling in temperate agricultural systems. In *Advances in Agronomy*. pp. 55–89.

Finnegan, E.J., Genger, R.K., Peacock, W.J. and Dennis, E.S. (1998) Dna methylation in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 223–247.

Flæte, N.E.S., Hollung, K., Ruud, L., Sogn, T., Færgestad, E.M., Skarpeid, H.J., Magnus, E.M. and Uhlen, A.K. (2005) Combined nitrogen and sulphur fertilisation and its effect on wheat quality and protein composition measured by SE-FPLC and proteomics. *J. Cereal Sci.*, **41**, 357–369.

Fox, S.E., Geniza, M., Hanumappa, M., et al. (2014) De novo transcriptome assembly and analyses of gene expression during photomorphogenesis in diploid wheat *Triticum monococcum* G. K. Pandey, ed. *PLoS One*, **9**, e96855.

Goldberg, T., Hecht, M., Hamp, T., et al. (2014) LocTree3 prediction of localization. *Nucleic Acids Res.*, **42**, W350–W355.

Grafi, G., Zemach, A. and Pitto, L. (2007) Methyl-CpG-binding domain (MBD) proteins in plants. *Biochim. Biophys. Acta - Gene Struct. Expr.*, **1769**, 287–294.

Granvogl, M., Wieser, H., Koehler, P., Tucher, S. Von and Schieberle, P. (2007) Influence of sulfur fertilization on the amounts of free amino acids in wheat. Correlation with baking properties as well as with 3-aminopropionamide and acrylamide generation during baking. *J. Agric. Food Chem.*, **55**, 4271–4277.

Harshman, S.W., Young, N.L., Parthun, M.R. and Freitas, M.A. (2013) H1 histones: current perspectives and challenges. *Nucleic Acids Res.*, **41**, 9593–9609.

Hidalgo, A. and Brandolini, A. (2014) Nutritional properties of einkorn wheat (*Triticum monococcum* L.). *J. Sci. Food Agric.*, **94**, 601–612.

Hidalgo, A., Brandolini, A., Pompei, C. and Piscozzi, R. (2006) Carotenoids and tocols of einkorn wheat (*Triticum monococcum* ssp. *monococcum* L.). *J. Cereal Sci.*, **44**, 182–193.

Hill, K., Horváth-Szanics, E., Hajós, G. and Kiss, É.

(2008) Surface and interfacial properties of water-soluble wheat proteins. *Colloids Surfaces A Physicochem. Eng. Asp.*, **319**, 180–187.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, **35**, W585–W587.

Husson, F., Josse, J., Le, S., Mazet, J. and & Husson, M.F. (2015) Package "FactoMineR".

Linder, P. (2006) Dead-box proteins: a family affair-active and passive players in RNP-remodeling. *Nucleic Acids Res.*, **34**, 4168–4180.

Liu, L., Zhang, Z., Mei, Q. and Chen, M. (2013) PSI: A comprehensive and integrative approach for accurate plant subcellular localization prediction J. L. Heazlewood, ed. *PLoS One*, **8**, e75826.

Locatelli, S., Piatti, P., Motto, M. and Rossi, V. (2009) Chromatin and DNA modifications in the Opaque2-mediated regulation of gene transcription during maize endosperm development. *Plant Cell*, **21**, 1410–1427.

Marcussen, T., Sandve, S.R., Heier, L., et al. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, **345**, 1250092–1250092.

Marion, D., Nicolas, Y., Popineau, Y., Branlard, G. and Landry, J. (1994) *A new and improved sequential extraction procedure of wheat proteins.* Molecular and functional aspects (p.47), ed., Presented at Wheat kernel proteins, Viterbo, ITA (1994-09-28 - 1994-09-30).

Martre, P., Porter, J.R., Jamieson, P.D. and Triboï (2003) Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. *Plant Physiol.*, **133**, 1959–1967.

Merkle, T. (2011) Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant Cell Rep.*, **30**, 153–176.

Mossé, J., Huet, J.C. and Baudet, J. (1985) The amino acid composition of wheat grain as a function of nitrogen content. *J. Cereal Sci.*, **3**, 115–130.

Nakamura, T., Yamaguchi, Y. and Sano, H. (1999) Four rice genes encoding cysteine synthase: isolation and differential responses to sulfur, nitrogen and light. *Gene*, **229**, 155–161.

Noctor, G. and Foyer, C.H. (1998) Ascorbate and Glutathione: keeping active oxygen under control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 249–279.

Oury, F.-X. and Godin, C. (2007) Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favourable genotypes? *Euphytica*, **157**, 45–57.

Petrovská, B., Šebela, M. and Doležel, J. (2015) Inside a plant nucleus: discovering the proteins. *J. Exp. Bot.*, **66**, 1627–1640.

Plessis, A., Ravel, C., Bordes, J., Balfourier, F. and Martre, P. (2013) Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. *J. Exp. Bot.*, **64**, 3627–3644.

R Core Team (2015) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Radchuk, V. V., Sreenivasulu, N., Radchuk, R.I., Wobus, U. and Weschke, W. (2005) The methylation cycle and its possible functions in barley endosperm development. *Plant Mol. Biol.*, **59**, 289–307.

Rea, M., Zheng, W., Chen, M., Braud, C., Bhangu, D., Rognan, T.N. and Xiao, W. (2012) Histone H1 affects gene imprinting and DNA methylation in *Arabidopsis*. *Plant J.*, **71**, 776–786.

Reinbold, J., Rychlik, M., Asam, S., Wieser, H. and Koehler, P. (2008) Concentrations of total glutathione and cysteine in wheat flour as affected by sulfur deficiency and correlation to quality parameters. *J. Agric. Food Chem.*, **56**, 6844–6850.

Rhazi, L., Cazalis, R., Lemelin, E. and Aussenac, T. (2003) Changes in the glutathione thiol–disulfide status during wheat grain development. *Plant Physiol. Biochem.*, **41**, 895–902.

Rossi, V., Motto, M. and Pellegrini, L. (1997) Analysis of the methylation pattern of the maize Opaque-2 (O2) promoter and in vitro binding studies indicate that the O2 B-Zip protein and other endosperm factors can bind to methylated target sequences. *J. Biol. Chem.*, **272**, 13758–13765.

Rubio-Somoza, I., Martinez, M., Abraham, Z., Diaz, I. and Carbonero, P. (2006) Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. *Plant J.*, **47**, 269–281.

Shewry, P.R. and Halford, N.G. (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.*, **53**, 947–958.

Shewry, P.R., Mitchell, R. a. C., Tosi, P., et al. (2012) An integrated study of grain development of wheat (cv. Hereward). *J. Cereal Sci.*, **56**, 21–30.

Shewry, P.R., Tatham, A.S. and Halford, N.G. (2001) Nutritional control of storage protein synthesis in developing grain of wheat and barley. *Plant Growth Regul.*, **34**, 105–111.

Shewry, P.R., Tatham, A.S. and Lazzeri, P. (1997) Biotechnology of wheat quality. *J. Sci. Food Agric.*, **73**, 397–406.

Simmonds, N.W. (1995) The relation between yield and protein in cereal grain. *J. Sci. Food Agric.*, **67**, 309–315.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Sørensen, M.B., Müller, M., Skerritt, J. and Simpson, D. (1996) Hordein promoter methylation and transcriptional activity in wild-type and mutant barley endosperm. *Mol. Gen. Genet.*, **250**, 750–

760.

Steinfurth, D., Zörb, C., Braukmann, F. and Mühling, K.H. (2012) Time-dependent distribution of sulphur, sulphate and glutathione in wheat tissues and grain as affected by three sulphur fertilization levels and late S fertilization. *J. Plant Physiol.*, **169**, 72–77.

Sturaro, M. and Viotti, A. (2001) Methylation of the Opaque2 box in zein genes is parent-dependent and affects O2 DNA binding activity in vitro. *Plant Mol. Biol.*, **46**, 549–560.

Tamura, K. and Hara-Nishimura, I. (2013) The molecular architecture of the plant nuclear pore complex. *J. Exp. Bot.*, **64**, 823–832.

Triboï, E., Martre, P. and Triboï-Blondel, A.-M. (2003) Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. *J. Exp. Bot.*, **54**, 1731–1742.

Tsuda, K., Tsuji, T., Hirose, S. and Yamazaki, K. (2004) Three *Arabidopsis* MBF1 homologs with distinct expression profiles play roles as transcriptional co-activators. *Plant Cell Physiol.*, **45**, 225–231.

Valot, B., Langella, O., Nano, E. and Zivy, M. (2011) MassChroQ: A versatile tool for mass spectrometry quantification. *Proteomics*, **11**, 3572–3577.

Vanyushin, B.F. and Ashapkin, V. V (2011) DNA methylation in higher plants: Past, present and future. *Biochim. Biophys. Acta*, **1809**, 360–368.

Verdier, J. and Thompson, R.D. (2008) Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.*, **49**, 1263–1271.

Vincent, J., Martre, P., Gouriou, B., Ravel, C., Dai, Z., Petit, J.-M. and Pailloux, M. (2015) RulNet: A web-oriented platform for regulatory network inference, application to wheat –omics data. *PLoS One*, **10**, e0127127.

Weis, B.L., Palm, D., Missbach, S., Bohnsack, M.T. and Schleiff, E. (2015) atBRX1-1 and atBRX1-2 are involved in an alternative rRNA processing pathway in *Arabidopsis thaliana*. *RNA*, **21**, 415–425.

Wierzbicki, A.T. and Jerzmanowski, A. (2005) Suppression of histone H1 genes in *Arabidopsis* results in heritable developmental defects and stochastic changes in DNA methylation. *Genetics*, **169**, 997–1008.

Wieser, H. (2007) Chemistry of gluten proteins. *Food Microbiol.*, **24**, 115–119.

Wieser, H., Antes, S. and Seilmeier, W. (1998) Quantitative determination of gluten protein types in wheat flour by reversed-phase high-performance liquid chromatography. *Cereal Chem.*, **75**, 644–650.

Wieser, H., Gutser, R. and Tucher, S. von (2004) Influence of sulphur fertilisation on quantities and proportions of gluten protein types in wheat flour. *J. Cereal Sci.*, **40**, 239–244.

Wieser, H. and Seilmeier, W. (1998) The influence of nitrogen fertilisation on quantities and proportions of different protein types in wheat flour. *J. Sci. Food Agric.*, **76**, 49–55.

Wrigley, C., Cros, D., Archer, M., Downie, P. and Roxburgh, C. (1980) The sulfur content of wheat endosperm proteins and its relevance to grain quality. *Aust. J. Plant Physiol.*, **7**, 755–766.

Yin, X. and Komatsu, S. (2015) Quantitative proteomics of nuclear phosphoproteins in the root tip of soybean during the initial stages of flooding stress. *J. Proteomics*, **119**, 183–195.

Zhao, F., Hawkesford, M. and McGrath, S. (1999) Sulphur assimilation and effects on yield and quality of wheat. *J. Cereal Sci.*, **30**, 1–17.

Zörb, C., Grover, C., Steinfurth, D. and Hermann Mühling, K. (2010) Quantitative proteome analysis of wheat gluten as influenced by N and S nutrition. *Plant Soil*, **327**, 225–234.

Supplementary information

Method S1. Semantics used to infer regulatory network from proteomics data.

Query 1: Discover rules between nutrition comparisons and relative abundance of proteins.

```
FINDRULES
SCOPE t1 IN Data_proteomique_RulNet_scaled
HAVING high: t1.ATT > 0.5 OVER ALL MINUS Stage, code_treatment, replicate, treatment, ST_treat
AND low: t1.ATT < -0.5 OVER ALL MINUS Stage, code_treatment, replicate, treatment, ST_treat
AND Nitrogen: (t1.ATT = 'T1' OR t1.ATT = 'T3') OVER code_treatment
AND NoNitrogen: (t1.ATT = 'C' OR t1.ATT = 'T2') OVER code_treatment
AND NoSulfur: (t1.ATT = 'C' OR t1.ATT = 'T1') OVER code_treatment
AND Sulfur: (t1.ATT = 'T2' OR t1.ATT = 'T3') OVER code_treatment
AND 300: t1.ATT = 300 OVER Stage
AND 400: t1.ATT = 400 OVER Stage
AND 500: t1.ATT = 500 OVER Stage
```

Query 2: Discover rules between co-accumulated proteins

```
FINDRULES
SCOPE t1 IN Data_proteomique_RulNet_scaled, t2 IN Data_proteomique_RulNet_scaled
WHERE t1.Stage = t2.Stage AND t1.code_treatment < t2.code_treatment
HAVING higher: t2.ATT-t1.ATT > 0.5 OVER ALL MINUS Stage, code_treatment, replicate, treatment, ST_treat
AND lower: t1.ATT-t2.ATT > 0.5 OVER ALL MINUS Stage, code_treatment, replicate, treatment, ST_treat
AND C_vs_T1: t1.ATT='C' AND t2.ATT = 'T1' OVER code_treatment
AND C_vs_T2: t1.ATT='C' AND t2.ATT = 'T2' OVER code_treatment
AND C_vs_T3: t1.ATT='C' AND t2.ATT = 'T3' OVER code_treatment
AND T1_vs_T2: t1.ATT='T1' AND t2.ATT = 'T2' OVER code_treatment
AND T1_vs_T3: t1.ATT='T1' AND t2.ATT = 'T3' OVER code_treatment
AND T2_vs_T3: t1.ATT='T2' AND t2.ATT = 'T3' OVER code_treatment
```

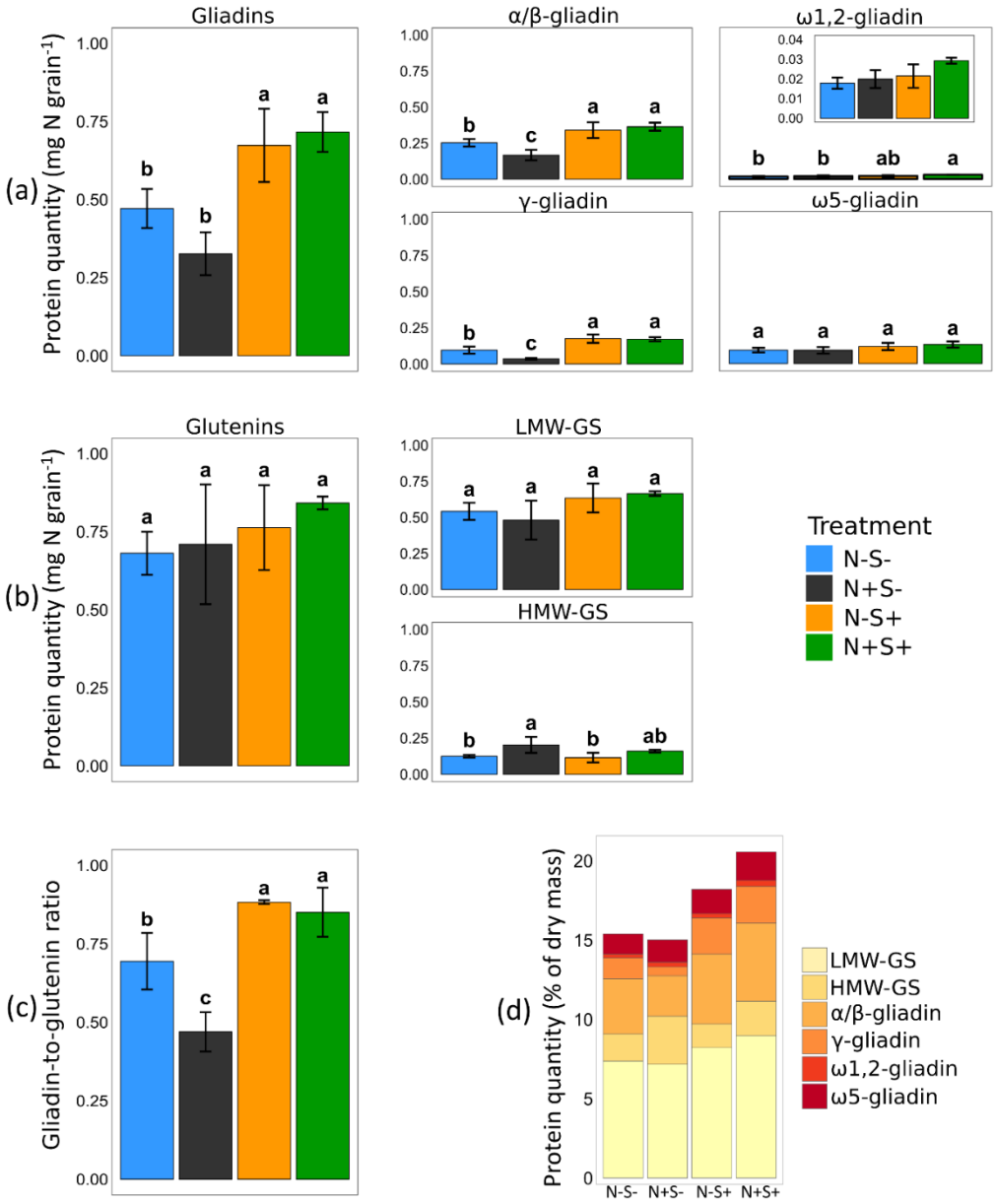


Figure S1. Effects of N and S supply on einkorn grain storage protein (GSP) composition at maturity. The four treatments were N-S- (0 mM N and 0 mM S), N+S- (6 mM N and 0 mM S), N-S+ (0.5 mM N and 2 mM S) and N+S+ (6 mM N and 2 mM S). (a) Quantity per grain of total gliadins, α/β-, γ-, ω1,2-, and ω5-gliadin classes. (b) Quantity per grain of total glutenins, LMW-GS, and HMW-GS. (c) Gliadin-to-glutenin ratio. (d) Grain storage protein composition expressed as percentage of grain dry mass. In (a), (b) and (c) data are mean ± 1 s.d. for $n = 4$ independent replicates and different letters above the error bars indicate statistical differences at $P < 0.05$.

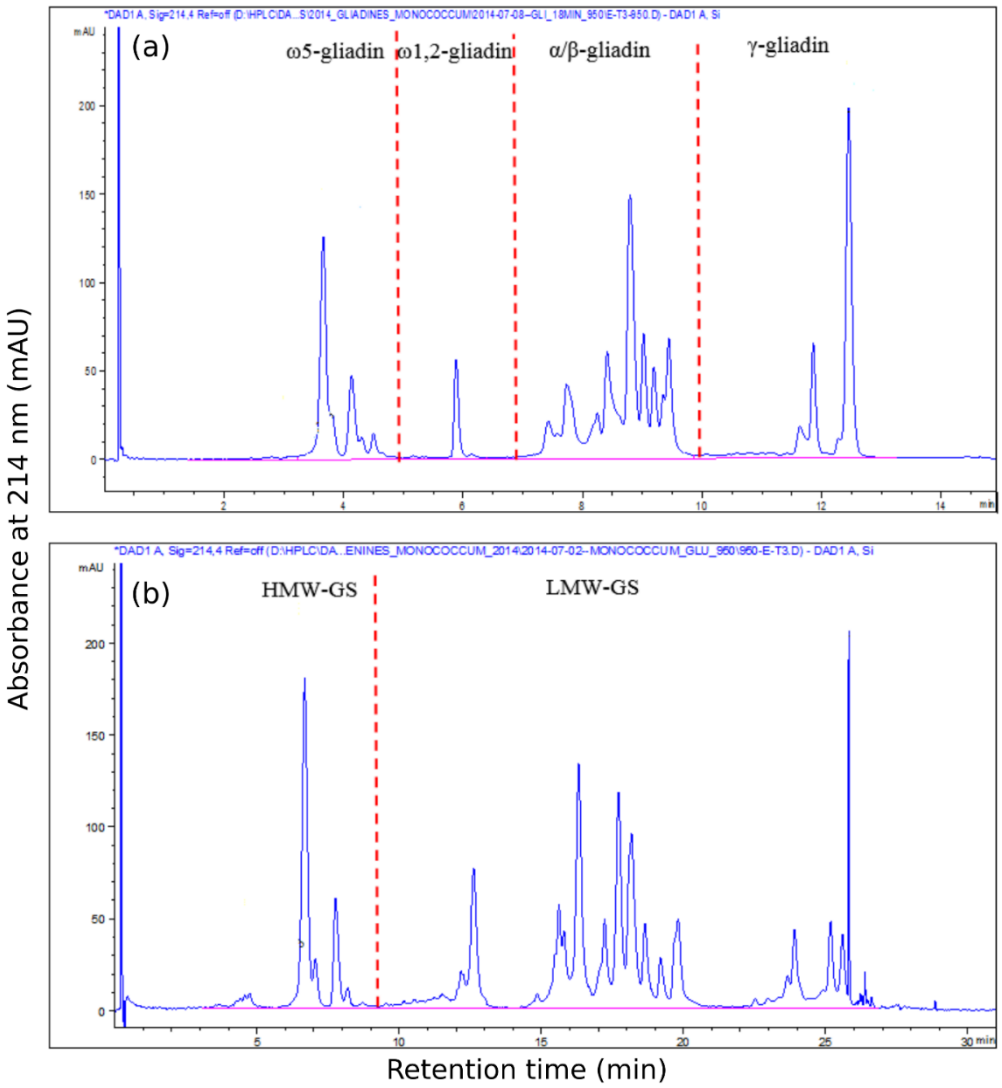


Figure S2. HPLC chromatograms of gliadin (a) and glutenin (b) in mature einkorn grain. Red dotted lines indicate separations between ω5-, ω1,2-, α/β- and γ-gliadin classes (a) and between HMW-GS and LMW-GS subunits (b).

Table S1. Quantity of dry mass, N, S and GSP per grain at maturity. Different letters within a row indicate significant differences ($P < 0.05$) between treatments. Data are mean ± 1 s.e. for $n = 4$ independent replicates.

	Treatments				***P*-value**
Variables	**N-S-**	**N+S-**	**N-S+**	**N+S+**	
Grain dry mass (mg DM grain^{-1})	41.8 ± 1.17^{a}	38.3 ± 5.07^{a}	43.9 ± 1.15^{a}	42.2 ± 0.72^{a}	0.523
Grain protein concentration (%)	19.3 ± 1.22^{b}	27.0 ± 1.96^{a}	23.6 ± 1.28^{ab}	27.4 ± 0.45^{a}	0.004
Grain S concentration (%)	0.19 ± 0.02^{b}	0.19 ± 0.02^{b}	0.29 ± 0.02^{a}	0.29 ± 0.01^{a}	< 0.001
N- to to-S ratio (-)	18.1 ± 0.78^{b}	24.8 ± 1.07^{a}	14.1 ± 0.23^{c}	16.5 ± 0.29^{bc}	< 0.001
Grain N (mg N grain^{-1})	1.42 ± 0.08^{b}	1.76 ± 0.15^{ab}	1.82 ± 0.10^{a}	2.03 ± 0.02^{a}	0.006
Grain S (mg S grain^{-1})	0.08 ± 0.01^{b}	0.07 ± 0.01^{b}	0.13 ± 0.01^{a}	0.12 ± 0.01^{a}	< 0.001

Table S2. Estimated maximum rate and duration of accumulation of grain dry mass, total N, total S and GSPs. Data shown in Figures 1 and 2 were fitted against thermal time using eqn. (1). Data are mean ± 1 s.e. *P*-values of the estimated maximum rate and duration are given in square brackets.

Variables	Treatments	Duration (°Cd)	Maximum rate (mg grain^{-1} [°Cd]$^{-1}$)
Dry mass	N-S-	590 ± 49 [7×10^{-11}]	(1.37 ± 0.25) × 10^{-1} [2×10^{-6}]
	N+S-	600 ± 63 [1×10^{-9}]	(1.12 ± 0.24) × 10^{-1} [7×10^{-6}]
	N- S+	633 ± 58 [1×10^{-10}]	(1.2 ± 0.21) × 10^{-1} [1×10^{-6}]
	N+S+	710 ± 77 [8×10^{-10}]	(0.92 ± 0.17) × 10^{-1} [1×10^{-6}]
N	N-S-	550 ± 46 [2×10^{-10}]	(5.87 ± 1.19) × 10^{-3} [1×10^{-5}]
	N+S-	702 ± 67 [3×10^{-10}]	(4.13 ± 0.68) × 10^{-3} [7×10^{-7}]
	N- S+	688 ± 63 [2×10^{-10}]	(4.42 ± 0.71) × 10^{-3} [4×10^{-7}]
	N+S+	691 ± 53 [2×10^{-11}]	(5.13 ± 0.71) × 10^{-3} [8×10^{-8}]
S	N-S-	502 ± 79 [1×10^{-6}]	(3.65 ± 1.5) × 10^{-4} [6×10^{-3}]
	N+S-	2015 ± 3674 [8×10^{-4}]	(0.55 ± 0.21) × 10^{-4} [3×10^{-2}]
	N- S+	798 ± 128 [2×10^{-7}]	(2.4 ± 0.56) × 10^{-4} [4×10^{-5}]
	N+S+	753 ± 119 [2×10^{-7}]	(2.5 ± 0.61) × 10^{-4} [8×10^{-5}]
LMW-GS	N-S-	595 ± 58 [3×10^{-7}]	(2.15 ± 0.54) × 10^{-3} [2×10^{-3}]
	N+S-	725 ± 118 [5×10^{-5}]	(1.11 ± 0.31) × 10^{-3} [3×10^{-3}]
	N- S+	705 ± 76 [8×10^{-7}]	(1.63 ± 0.33) × 10^{-3} [4×10^{-4}]
	N+S+	717 ± 74 [5×10^{-7}]	(1.68 ± 0.32) × 10^{-3} [2×10^{-4}]
HMW-GS	N-S-	615 ± 44 [8×10^{-9}]	(6.02 ± 1.36) × 10^{-4} [8×10^{-4}]
	N+S-	624 ± 31 [1×10^{-10}]	(8.39 ± 1.2) × 10^{-4} [1×10^{-5}]
	N- S+	946 ± 132 [1×10^{-5}]	(2.32 ± 0.47) × 10^{-4} [4×10^{-4}]
	N+S+	818 ± 67 [4×10^{-8}]	(3.91 ± 0.6) × 10^{-4} [3×10^{-5}]
α/β-gliadin	N-S-	730 ± 150 [4×10^{-4}]	(5.5 ± 1.87) × 10^{-4} [1×10^{-2}]
	N+S-	772 ± 266 [1×10^{-2}]	(3.42 ± 1.74) × 10^{-4} [7×10^{-2}]
	N- S+	1049 ± 237 [8×10^{-4}]	(4.91 ± 1.12) × 10^{-4} [9×10^{-4}]
	N+S+	898 ± 136 [3×10^{-5}]	(6.55 ± 1.34) × 10^{-4} [4×10^{-4}]
γ-gliadin	N-S-	928 ± 421 [5×10^{-2}]	(1.33 ± 0.69) × 10^{-4} [8×10^{-2}]
	N+S-	1666 ± 14680 [9×10^{-1}]	(0.21 ± 1.13) × 10^{-4} [9×10^{-1}]
	N- S+	1101 ± 250 [9×10^{-4}]	(2.49 ± 0.52) × 10^{-4} [5×10^{-4}]
	N+S+	1000 ± 188 [2×10^{-4}]	(2.73 ± 0.58) × 10^{-4} [5×10^{-4}]
ω1,2-gliadin	N-S-	707 ± 109 [3×10^{-5}]	(4.41 ± 1.24) × 10^{-5} [4×10^{-3}]
	N+S-	684 ± 74 [8×10^{-7}]	(6.31 ± 1.4) × 10^{-5} [7×10^{-4}]
	N- S+	1292 ± 383 [6×10^{-3}]	(2.89 ± 0.52) × 10^{-5} [1×10^{-4}]
	N+S+	926 ± 98 [6×10^{-7}]	(5.56 ± 0.81) × 10^{-5} [2×10^{-5}]
ω5-gliadin	N-S-	736 ± 107 [2×10^{-5}]	(2.25 ± 0.58) × 10^{-4} [2×10^{-3}]
	N+S-	698 ± 78 [1×10^{-6}]	(2.97 ± 0.69) × 10^{-4} [1×10^{-3}]
	N- S+	1033 ± 175 [7×10^{-5}]	(1.9 ± 0.35) × 10^{-4} [2×10^{-4}]
	N+S+	808 ± 79 [3×10^{-7}]	(3.06 ± 0.5) × 10^{-4} [5×10^{-5}]

Tables S3 and S4 are not necessary to understand the PhD manuscript since they correspond to quantitative data for GSP, albumin-globulin and nuclear proteins. Thus they have not been included in this manuscript.

Table S5. Summary of the 209 proteins significantly impacted by N and S supply. Column 'Type' mentions if protein is part of the Grain Storage Protein (GSP), nuclear (nucl) or albumin-globulin (alg) fraction. Column 'Code' provides codes used in figures to qualify proteins, for nuclear and albumin-globulin proteins it corresponds to their Uniprot accession number. Then cluster number and connectivity of these proteins in network analyses are provided. In columns related to the nutrition network, sign '+' or '-' indicate that the protein was present in network, a '+' indicating a higher amount in the first treatment and a '-' indicating a lower amount in the first treatment. The column named 'connectivity' indicates the number of edges relied to the protein, and thus reflects association number.

Type	Code	Description	Cluster	Nutrition network						Network co-accumulation
				N+S- N-S-	N-S+ N-S-	N+S+ N-S-	N-S+ N+S-	N+S+ N+S-	N+S+ N-S+	Connectivity
GSP	HMW	HMW-GS	2	+			-	-		5
GSP	LMW	LMW-GS	4			+				5
GSP	α/β	α/β-gliadin	4			+		+		
GSP	γ	γ-gliadin	3			+	+	+		
GSP	ω1.2	ω1.2	2			+	-		+	4
GSP	ω5	ω5-gliadin	4						+	5
Nucl	B6RRA6	Puroindoline A	3	-			+	+		
Nucl	D8L9M8	RNA recognition domain containing protein,expressed	4			+		+		
Nucl	E2F3W4	40S Ribosomal protein S15a	4					+		2
Nucl	F4Y595	Heat shock protein 9	4							
Nucl	M7YCG2	Importin subunit alpha	4					+		
Nucl	M7YGL9	Alanine aminotransferase 2	4			+				4
Nucl	M7YRR2	Actin-depolymerizing factor 4	4							4
Nucl	M7YRS8	Histone H1	1				-			
Nucl	M7YRT1	Crooked neck-like protein 1	4						+	
Nucl	M7YWU3	Heat shock cognate 7 kDa protein 1	4					+		
Nucl	M7YZD9	60S Ribosomal protein L18	1			-				4
Nucl	M7ZFZ3	DNA-directed RNA polymerases I, II, and III subunit RPABC1	4							15
Nucl	M7ZGF0	dna-binding protein	2				-			
Nucl	M7ZJH4	14-3-3-like protein B	4							3
Nucl	M7ZPU0	peter pan-like protein	4					+	+	6
Nucl	M7ZPU9	DEAD-box ATP-dependent RNA helicase 27	4			+				11
Nucl	M7ZXC1	phosphorylase	4							18
Nucl	M8A0B6	Histone H1	2				-			
Nucl	M8A0L6	coiled-coil domain-containing protein 12	4							2
Nucl	M8A407	60S acidic Ribosomal protein P2B	3		+		+			
Nucl	M8A553	60S Ribosomal protein L23a	4					+	+	
Nucl	M8A571	Elongation factor 2	4					+		
Nucl	M8AAX5	Multiprotein-bridging factor 1a	4							12
Nucl	M8AL83	40S Ribosomal protein S11	4							9
Nucl	M8AVG0	Serine/arginine-rich splicing factor 4	4							
Nucl	P12783	phosphoglycerate kinase, cytolic	4							5
Nucl	P93616	Poly(A)-binding protein	4							
Nucl	Q03387	Eukaryotic translation initiation factor isoform 4G-1	4					+		6
Nucl	Q3S4I0	Eukaryotic translation initiation factor 5A2	4							
Nucl	Q9XHL9	Histone H1 WH1B.1	2							
Nucl	T1LTD9	tyrosine-protein phosphatase	2							3
Nucl	T1MN05	fam10 family protein at4g22670-like	4							
Nucl	T1N8M0	probable aldo-keto reductase 1	4							14
Nucl	W5AA91	Importin subunit alpha	4							
Nucl	W5AGD5	trihelix transcription factor	4	+						3
Nucl	W5ANL4	protein rrp5 homolog	4							10
Nucl	W5AP12	DNA-directed RNA polymerase	4							
Nucl	W5B117	guanine nucleotide-binding 3 homolog	4			+				
Nucl	W5BW98	protein rcc2	4		+					
Nucl	W5CME5	60s ribosomal protein l30	1			-				1
Nucl	W5DPT1	Spastin	2							1
Nucl	W5EU71	formate-tetrahydrofolate ligase	4							8

Nucl	W5F910	dead-box atp-dependent rna helicase 28	4					+	+	2
Nucl	W5FES4	brain acid soluble protein 1	3				+			
Nucl	W5G441	nucleolar complex protein 3 homolog	4							15
Nucl	W5G6L7	betaine aldehyde dehydrogenase	4			+				14
Nucl	W5HA67	hypothetical protein F775_28656	3							1
Nucl	W5HDD8	sucrose synthase	3					+		
Alg	A1YUM8	Limit dextrinase type starch debranching enzyme	3	-			+	+		
Alg	A9U8G4	Alcohol dehydrogenase ADH1A	3	-			+	+		8
Alg	A9YPI3	Puroindoline A	3	-		-	+			
Alg	B2CGM5	Triticin	3					+		
Alg	B2ZGG4	GTPase SAR1	1							
Alg	B3TZE7	Subtilisin protease (Fragment)	2	+						
Alg	B6DU63	Zeta-carotene desaturase	2	+			-	-		
Alg	B6UKS0	Gamma-gliadin	3				+			
Alg	B8XU49	Gamma gliadin (Fragment)	1							
Alg	C0KTA6	Fructose-bisphosphate aldolase	4							3
Alg	C5J3R4	Trypsin inhibitor	3	-			+	+		1
Alg	D2KZ08	Aminotransferase	2					-		1
Alg	D3KVP3	Purple acid phosphatase	1			-				7
Alg	D6QY49	Omega-gliadin (Fragment)	2				-			4
Alg	D8L9B3	Protein disulfide isomerase family protein 4-1	1							4
Alg	D8L9M8	RNA recognition domain containing protein,expressed	2							1
Alg	F6LAX4	Protein phosphatase 2A structural subunit	2							
Alg	G0YVZ4	High molecular weight glutenin 1Ax2.1	2							4
Alg	G3CCE7	Starch branching enzyme IIa	1			-				
Alg	H9NAV2	Superoxide dismutase [Cu-Zn] (Fragment)	1							
Alg	M7Y8M9	Translocon-associated protein subunit beta	1							
Alg	M7YBY6	Basic endochitinase C	3							8
Alg	M7YCG2	Importin subunit alpha	2						+	
Alg	M7YE46	Vicilin-like antimicrobial peptides 2-2	3							11
Alg	M7YIZ6	Xylulose kinase	1	-	-	-				1
Alg	M7YM33	Isocitrate dehydrogenase [NAD] catalytic subunit 5, mitochondrial	2							
Alg	M7YMQ8	GrpE protein homolog	4							1
Alg	M7YMT4	Putative aconitate hydratase, cytoplasmic	3				+			
Alg	M7YPB1	Glutaredoxin-C6	4			+			+	1
Alg	M7YQA5	Aminopeptidase N	4							3
Alg	M7YQW7	Acyl-[acyl-carrier-protein] desaturase 2, chloroplastic	4							
Alg	M7YT47	zinc finger protein	1							
Alg	M7YUM7	Intracellular protease 1	3	-			+			
Alg	M7YUQ6	Aspartic proteinase oryzasin-1	1							7
Alg	M7YVM9	Chitinase 1	4			+				12
Alg	M7Z025	Ankyrin repeat domain-containing protein 2	3							10
Alg	M7Z0D7	DNA damage-inducible protein 1	1							
Alg	M7Z2B3	transmembrane 214-B	1			-				
Alg	M7Z2D6	amino acid selective channel protein	1							
Alg	M7Z2J3	Putative NADP-dependent oxidoreductase P1	4			+				10
Alg	M7Z3I6	2-oxoglutarate dehydrogenase, mitochondrial	4							2
Alg	M7Z3P0	Peroxisomal multifunctional enzyme type 2	3	-						
Alg	M7Z3U2	non-specific lipid transfer GPI-anchored 2-like	3	-						
Alg	M7Z534	Glutaredoxin-C8	3				+			
Alg	M7Z7R6	isoflavone reductase homolog irl-like	2	+			-	-		
Alg	M7Z8W1	Cytochrome b5	1							1
Alg	M7ZA33	Serpin-ZX	1					-		4
Alg	M7ZA82	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha	2		-		-			
Alg	M7ZDG7	Monothiol glutaredoxin-S11	4			+				
Alg	M7ZF09	Endogenous alpha-amylase/subtilisin inhibitor	4							10
Alg	M7ZGB1	Programmed cell death protein 4	2			+	-		+	
Alg	M7ZIQ2	Putative glutathione S-transferase GSTF1	3				+	+		
Alg	M7ZIY2	2'-deoxymugineic-acid 2'-dioxygenase	4							
Alg	M7ZJJ1	Peptide methionine sulfoxide reductase A2-1	2							1
Alg	M7ZKE7	glutathionyl-hydroquinone reductase -like	2	+			-	-		
Alg	M7ZMS4	Endoglucanase 11	3				+			
Alg	M7ZNK7	proton pump-interactor 1-like	1							
Alg	M7ZQM4	L-ascorbate peroxidase 1, cytosolic	4							15
Alg	M7ZRM5	Programmed cell death protein 4	2				-			4
Alg	M7ZT31	Gamma-hordein-3	3	-	+		+	+	-	
Alg	M7ZVP6	Bifunctional aspartokinase/homoserine dehydrogenase 2, chloroplastic	3							
Alg	M7ZXX2	Pyruvate, phosphate dikinase 1, chloroplastic	3	-		-				

Alg	M7ZY63	Coatomer subunit gamma	2		-		-		+	
Alg	M7ZZM5	Adenosine kinase 2	4							9
Alg	M8A0T7	Sucrose synthase 2	1							
Alg	M8A1S2	Trypsin/alpha-amylase inhibitor CMX1/CMX3	3	-			+	+	-	
Alg	M8A1U1	UDP-glucuronic acid decarboxylase 1	4							5
Alg	M8A2G0	ADP,ATP carrier protein, mitochondrial	1							
Alg	M8A2R3	Uncharacterized protein	1							
Alg	M8A3G4	Cell division protein ftsZ-like protein 2-1, chloroplastic	2				-		+	
Alg	M8A710	Clathrin heavy chain	2				-		+	
Alg	M8A7I9	Chitinase 2	4							13
Alg	M8A8D6	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	2	+						1
Alg	M8A8V1	hypothetical protein TRIUR3_12434	1							1
Alg	M8AMK8	Actin-7	1		-		-			4
Alg	M8APL9	DEAD-box ATP-dependent RNA helicase 37	2				-			
Alg	M8AQI0	Inositol-3-phosphate synthase	4							
Alg	M8AQI5	Avenin-3	3	-			+	+		
Alg	M8AS42	Putative bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase E1	3	-			+	+		
Alg	M8ATK6	Protein argonaute 4A	3							
Alg	M8AUX2	caleosin 2	1							
Alg	M8B5G5	Beta-amylase	3	-			+	+		
Alg	O04074	Starch branching enzyme 1	1							7
Alg	P12299	Glucose-1-phosphate adenylyltransferase large subunit, chloroplastic/amyloplastic	1							
Alg	P12782	Phosphoglycerate kinase, chloroplastic	4							14
Alg	P38076	Cysteine synthase	2	+			-	-		
Alg	P82901	Non-specific lipid-transfer protein 2P	3							2
Alg	Q1ALB3	Disproportionating enzyme	1							
Alg	Q24M29	Starch branching enzyme IIb	3	-			+	+		
Alg	Q3S832	Delta-1-pyrroline-5-carboxylate dehydrogenase	3		+		+			
Alg	Q58U44	Ribulose bisphosphate carboxylase large chain	4							12
Alg	Q6QFA1	Expansin EXPB3	3				+			
Alg	Q6RUJ1	Glutamine synthetase	4							1
Alg	Q7X9M4	Putative uncharacterized protein (Fragment)	3	-			+	+		
Alg	Q8RW02	Glutathione transferase	3				+	+		9
Alg	Q8RWR5	Beta-D-glucan exohydrolase	2	+				-	-	
Alg	Q8S4P7	Thaumatin-like protein	3					+		1
Alg	Q8W3V6	Low-molecular-weight glutenin subunit group 6 type IV	3	-			+	+		1
Alg	Q9FUU8	Starch branching enzyme 1	1							7
Alg	Q9M4Z1	Glucose-1-phosphate adenylyltransferase	1							
Alg	Q9ZRB0	Tubulin beta-3 chain	3					+		
Alg	T1LJ90	glutathione s-transferase 3	4							1
Alg	T1MV14	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	2			+			+	
Alg	T1N8M0	probable aldo-keto reductase 1-like	3					+		
Alg	T1NLZ4	Uncharacterized protein	3				+			
Alg	W4ZUH4	importin subunit beta-1-like	2							2
Alg	W4ZXS7	12s seed storage globulin 1-like	4							
Alg	W4ZY98	stromal 70 kda heat shock-related chloroplastic	1							
Alg	W5A2X8	universal stress protein a-like protein	1							1
Alg	W5A845	pyrophosphate-energized vacuolar membrane proton pump 1-like	1							
Alg	W5AA91	Importin subunit alpha	2	+		+	-		+	
Alg	W5AS47	clathrin light chain 2-like	2							
Alg	W5AUD8	glutamate--glyoxylate aminotransferase 2	2							
Alg	W5AUJ3	ECERIFERUM 26-like	4							
Alg	W5AWM0	methylthioribose kinase	2	+			-	-		
Alg	W5B0Q9	Xylose isomerase	3				+			
Alg	W5BD57	chaperone protein chloroplastic-like	2	+	-		-		+	
Alg	W5BEK6	Uncharacterized protein	2			+	-		+	
Alg	W5BKC3	udp-glucuronic acid decarboxylase 2-like	1							2
Alg	W5BLP9	Uncharacterized protein (Fragment)	2	+			-	-		
Alg	W5CCM5	carbonic anhydrase	3	-			+	+		
Alg	W5CFJ5	flower-specific gamma-thionin precursor	3				+		-	
Alg	W5CPX1	uridine 5 -monophosphate synthase-like	4							2
Alg	W5CR31	cationic peroxidase SPC4-like	3							11
Alg	W5D1Z6	pectin acetylesterase 5-like	1							15
Alg	W5DGP3	pyruvate decarboxylase	3				+	+		8
Alg	W5DLU4	glutamate-rich wd repeat-containing protein 1	4							
Alg	W5DP16	UDP-glucose 6-dehydrogenase	3	-			+			

Alg	W5DR83	gibberellin 20 oxidase 2	2	+	-		-	-		
Alg	W5DS49	rna recognition motif containing family protein	1							4
Alg	W5DWS0	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	1							
Alg	W5DZZ4	transport SEC23-like	2						+	
Alg	W5E0G4	vicilin-like antimicrobial peptides 2-2	3							11
Alg	W5E0G5	cop9 signalosome complex subunit 4	4							
Alg	W5E5E8	Alcohol dehydrogenase-like 2	3							9
Alg	W5EKP0	cop9 signalosome complex subunit 1	2							
Alg	W5EM56	phosphoenolpyruvate carboxykinase	3					+		
Alg	W5EQ81	cbs domain-containing protein	3							
Alg	W5F7H1	probable n-succinyldiaminopimelate aminotransferase	2				-	-		
Alg	W5F9D8	probable sarcosine oxidase	1							10
Alg	W5F9U4	probable sarcosine oxidase	1							7
Alg	W5FBU7	ripening-related 1	4			+		+		9
Alg	W5FEQ2	prohibitin- mitochondrial-like	1							
Alg	W5FF37	alpha,alpha-trehalose-phosphate synthase [UDP-forming] 6-like	2				-		+	
Alg	W5FHA6	Oleosin	1							14
Alg	W5FZQ1	rRNA N-glycosidase	1							5
Alg	W5G958	acetolactate synthase small subunit 2, chloroplastic-like	2		-					1
Alg	W5GC44	hyoscyamine 6-dioxygenase	3				+	+		
Alg	W5GWT6	aspartic proteinase oryzasin-1-like	1							
Alg	W5H829	protein kinase superfamily protein	2		-		-			
Alg	W5HAJ8	probable gamma-aminobutyrate transaminase 3, mitochondrial	2				-	-		
Alg	W5HFK6	Formate dehydrogenase, mitochondrial	1							
Alg	W5HXV0	fructokinase-2	3	-			+			1
Alg	W5HY12	Dolichyl-diphosphooligosaccharide-- glycosyltransferase 48 kDa subunit	1			-				
Alg	W5I774	sucrose synthase	4					+		8

3. CONCLUSIONS

Dans notre étude, les traitements nutritionnels appliqués à 200°Cj après floraison ont entrainé des quantités et des durées d'accumulation différentes des éléments N et S au sein du grain de *T. monococcum*. Cela a conduit à une modification du ratio N/S qui a été très impacté par un fort apport d'N sans apport conjoint de S. Dans cette condition, ce rapport N/S avoisinait 30, synonyme d'une sévère déficience en S (Randall *et al.*, 1981). A maturité du grain, la composition en PR a été nettement modifiée, avec une augmentation de l'accumulation des PR riches en S avec les traitements soufrés et une augmentation de celle des PR pauvres en S avec les traitements azotés. Ces résultats confirment ainsi ceux rapportés dans de nombreuses études chez le blé tendre (Wieser *et al.*, 2004 ; Zörb *et al.*, 2010 ; Dai *et al.*, 2015). L'apport de S a entrainé un taux élevé de gliadines, expliqué par une forte augmentation de la vitesse et de la durée d'accumulation des α/β- et γ-gliadines. Au final, le rapport gliadines/gluténines qui influence les propriétés du gluten a été clairement modifié par les différents traitements.

L'analyse des protéines nucléaires d'une part et des albumines globulines d'autre part a permis d'identifier et de quantifier de nombreuses protéines et ainsi d'obtenir une bonne vision du protéome du grain de *T. monococcum* en réponse à l'apport de N et de S. Même si cela n'a pas été discuté dans cette étude consacrée à la réponse du protéome à la nutrition, le nombre élevé de protéines nucléaires identifiées (1677) apporte de nouvelles informations sur les acteurs protéiques rencontrés dans le noyau de la cellule végétale.

Si la modification du ratio N/S a été responsable des changements de la composition en PR du grain, nous avons également supposé qu'il fût à l'origine des variations quantitatives observées pour de nombreuses albumines globulines et protéines nucléaires au cours du remplissage du grain. En effet, beaucoup de ces protéines ont présenté des variations d'abondance entre les deux conditions où l'N ou le S avaient été apportés seuls à la culture. Parmi les changements observés, l'augmentation de la quantité de protéines impliquées dans le métabolisme des acides aminés a été observée lorsqu'un fort apport d'N a été effectué, à l'inverse de plusieurs protéines liées au métabolisme des sucres.

Les protéines nucléaires ont été moins affectées par le changement de nutrition que les albumines globulines. La plupart de celles qui ont été impactées l'ont été en réponse à un apport élevé d'N et de S. Parmi elles, de nombreuses protéines impliquées dans la maturation des ARNm ont été retrouvées, synonyme d'une activation de la transcription dans cette condition. Les protéines de liaison à l'ADN ont quant à elle présenté des réponses très variées, suggérant une action isolée de ces protéines et spécifique d'un état nutritionnel du grain. Parmi ces réponses, une régulation par des mécanismes épigénétiques *via* la méthylation de l'ADN semblerait être activée suite à un fort apport de S. Les résultats observés au niveau des protéines impliquées dans le transport nucléo-cytoplasmique témoignent de l'importance des échanges d'ARN et de protéines entre le noyau et le cytoplasme. Même si peu de protéines nucléaires ont été impactées par la nutrition, le noyau semble occuper une place centrale dans l'ajustement de la composition du grain suite à différents apports de N et de S.

Les réseaux biologiques générés nous ont permis de visualiser les protéines réagissant de la même manière aux apports d'N et de S. Il est probable que ces groupes de protéines sont en partie responsables de l'ajustement du métabolisme cellulaire, à l'origine de la modification de

la composition du grain en PR. L'intégration de ces données de protéomique avec d'autres types de données -omiques semble importante pour détecter des acteurs centraux de la réponse à la nutrition, qui pourraient faire l'objet d'études plus approfondies.

Références

Dai, Z., Plessis, A., Vincent, J., et al. (2015) Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. *Plant J.*, **83**, 1–18.

Randall, P., Spencer, K. and Freney, J. (1981) Sulfur and nitrogen fertilizer effects on wheat. I. Concentrations of sulfur and nitrogen and the nitrogen to sulfur ratio in grain, in relation to the yield response. *Aust. J. Agric. Res.*, **32**, 203.

Wieser, H., Gutser, R. and Tucher, S. von (2004) Influence of sulphur fertilisation on quantities and proportions of gluten protein types in wheat flour. *J. Cereal Sci.*, **40**, 239–244.

Zörb, C., Grover, C., Steinfurth, D. and Hermann Mühling, K. (2010) Quantitative proteome analysis of wheat gluten as influenced by N and S nutrition. *Plant Soil*, **327**, 225–234.

CHAPITRE 5 :
La réponse intégrative du grain de *Triticum monococcum* à la nutrition azotée et soufrée

1. INTRODUCTION

Pour compléter l'étude des protéines nucléaires et des albumines globulines, des analyses du transcriptome et de plusieurs métabolites à partir du même matériel biologique ont été réalisées. Des études récentes chez le blé et le seigle montrant l'impact de la disponibilité en N et S sur l'expression de nombreux gènes du grain liés au métabolisme cellulaire et au transport des acides aminés ainsi que sur des régulateurs transcriptionnels ont motivé ces analyses (Dai *et al.*, 2015 ; Postles *et al.*, 2016).

Pour la partie transcriptomique, les ARN ont été quantifiés par séquençage d'ARN (RNA-Seq), entre 300 et 600°Cj après floraison pour observer les effets de la nutrition pendant le remplissage du grain, ainsi qu'à 100 et 200°Cj. Les métabolites étudiés correspondent à des sucres, des acides aminés, des acides organiques, le glutathion, ainsi que quelques dérivés aminés et nucléosides. Ils ont été analysés sur l'ensemble du développement du grain, de 100 à 950°Cj après floraison.

Après avoir analysé l'impact de l'N et du S sur l'expression des gènes d'une part et sur l'accumulation des métabolites d'autre part, nous avons intégré les données acquises à ces différentes échelles (transcrits, métabolites) en incluant les résultats de protéomique obtenus dans le chapitre précédent. L'objectif a ainsi été d'obtenir une vision globale de la réponse du grain de *T. monococcum* à des apports contrastés en N et S. La stratégie adoptée a été de générer des réseaux biologiques de co-expression/co-accumulation à partir de l'ensemble des variables répondant significativement à la nutrition. Un premier réseau a été généré afin de lier les ARN/protéines/métabolites ayant les mêmes profils d'abondance pendant le remplissage du grain. Le but a été de visualiser des groupes de variables intervenant potentiellement dans les mêmes voies de réponse. Puis comme dans le précédent chapitre, les variables répondant de la même façon aux différents traitements pendant le remplissage du grain ont été groupées au sein d'un second réseau. L'utilisation de ces deux types de réseaux, porteurs d'informations complémentaires, nous a semblé pertinente pour identifier les variables ayant des profils d'abondance similaires à la fois sur la cinétique étudiée et en réponse aux traitements azotés et soufrés.

La finalité de cette étude est de faire le lien entre le facteur environnemental étudié, la nutrition N et S et le phénotype observé, la composition du grain en PR, en considérant la réponse cellulaire comme un système dont les parties, qui correspondent aux différents types de données -omiques, sont fortement connectées (Figure 24). L'idée sous-jacente à ce travail est de comprendre cette réponse afin d'identifier les leviers d'action pour maitriser la composition protéique dans un contexte de limitation des engrais.

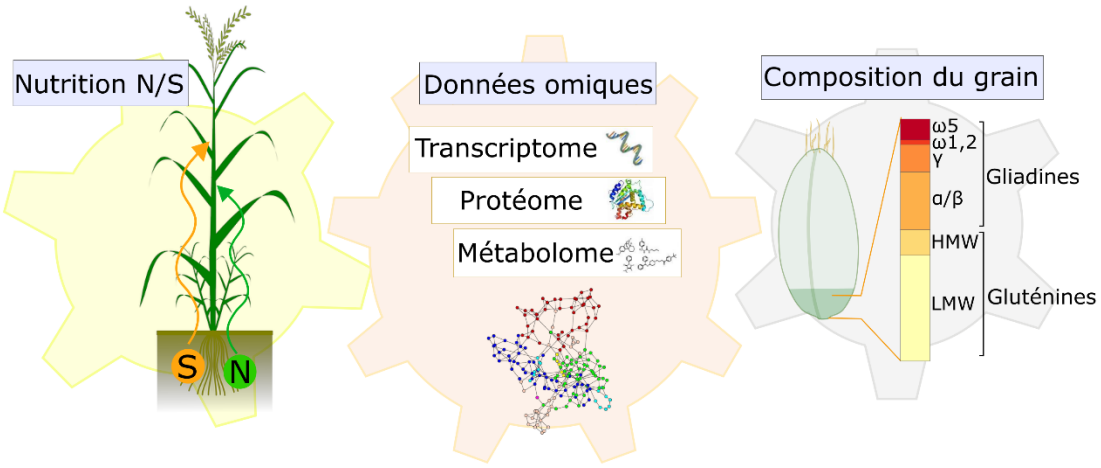


Figure 24. Représentation schématique de l'objectif du chapitre 5 : intégrer les réponses du grain de T. monococcum à l'apport de N et S aux différentes échelles mesurées (transcrits, protéines, métabolites) pour faire le lien entre la nutrition et la composition en protéines de réserve.

Références :

Dai, Z., Plessis, A., Vincent, J., et al. (2015) Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. *Plant J.*, **83**, 1–18.

Postles, J., Curtis, T.Y., Powers, S.J., Elmore, J.S., Mottram, D.S. and Halford, N.G. (2016) Changes in free amino acid concentration in rye grain in response to nitrogen and sulfur availability, and expression analysis of genes involved in asparagine metabolism. *Front. Plant Sci.*, **7**, 1–11.

2. ARTICLE 4 : Integrative response of the einkorn (*Triticum monococcum*) grain to nitrogen and sulphur nutrition

Titouan Bonnot[1], Victor Hatte[1], Emmanuelle Bancel[1], Mireille Dardervet[1], Philippe Leroy[1], Julie Boudet[1], Annick Moing[2], Yves Gibbon[2], Marie Pailloux[3], Pierre Martre[1] and Catherine Ravel[1]

[1] *UMR GDEC, INRA, Université Blaise Pascal, 5 chemin de Beaulieu, 63 177 Aubière, France*
[2] *INRA, UMR1332 Biologie du Fruit et Pathologie, Villenave d'Ornon F-33 882, France*
[3] *LIMOS, CNRS, Université Blaise Pascal, 63 173 Aubière, France*

(Article en préparation)

Integrative response of einkorn (*Triticum monococcum*) grain to nitrogen and sulphur nutrition

Titouan Bonnot[1], Victor Hatte[1], Emmanuelle Bancel[1], Mireille Dardervet[1], Philippe Leroy[1], Julie Boudet[1], Annick Moing[2], Yves Gibbon[2], Marie Pailloux[3], Catherine Ravel[1,*] and Pierre Martre[1, †]

[1] *UMR GDEC, INRA, Université Blaise Pascal, 5 chemin de Beaulieu, 63 177 Aubière, France*
[2] *INRA, UMR1332 Biologie du Fruit et Pathologie, Villenave d'Ornon F-33 882, France*
[3] *LIMOS, CNRS, Université Blaise Pascal, 63 173 Aubière, France*
[†]Present address: UMR LEPSE, INRA, Montpellier SupAgro, 34 060 Montpellier, France

Abstract: In wheat, one major challenge is to maintain grain quality while increasing yield and preserving the environment. To achieve this goal, it is essential to considerate the nitrogen (N) and sulfur (S) nutrition, which has long been known for its impact on the grain composition and particularly grain storage protein (GSP) composition. The identification and understanding of molecular mechanisms regulated by the nutrition during grain development is necessary. In this study four treatments with various N and S amounts were applied to an einkorn (*Triticum monococcum*) culture during grain development. Transcriptome and metabolite analyses performed at different time points and for the different treatments have highlighted the strong impact of S deficiency in the grain response to nutrition, conducting to the activation of amino acid and S transport and metabolism. An integrative view of this response was obtained by supplementing transcriptome and metabolome with proteomic data previously obtained from the same biological material. Major changes were observed when N-to-S ratio was disturbed in grain. Several transporters and genes involved in transcription process were revealed in network analysis and can have a central role in the grain adaptation to such changes. Finally, the large dataset obtained provides information for the identification of genes and pathways involved in the response to nutrition, which could allow a better control of the grain composition and thus its quality.

Keywords: einkorn, nitrogen, sulfur, grain, transcriptome, metabolites, proteins, co-expression network

INTRODUCTION

During grain development, storage compounds accumulate to form a nutrient stock used during germination process and first steps of seedling growth before acquisition of autotrophy. Among storages in wheat, grain storage proteins (GSP) provide N and S resources and account for 60 to 80% of the total protein amount in mature grain (Shewry and Halford, 2002). In addition to their essential role in seedling nutrition, quantity and composition of GSP in mature grain determine its end-use value for food. They are divided into two fractions: gliadins, which are monomeric proteins and glutenins, which form large macropolymers during grain maturation (Don *et al.*, 2006). Together they represent major constituents of gluten and determine the rheological properties of dough, with glutenins that are mainly responsible for viscoelastic properties and gliadins that confer extensibility to dough (Branlard *et al.*, 2001). Gliadins are subdivided into ω1,2-, ω5-, α/β- and γ-gliadin classes, and low-molecular-weight-subunits (LMW-GS) and high-molecular-weight-subunits (HMW-GS) subunits are distinguished among glutenins (Wieser, 2007). Proportion of these different classes/subunits determines grain quality.

N and S are two essential elements for plant nutrition, growth and accumulation of storages in grain. Indeed they represent two major constituents of proteins and other organic compounds and many studies have previously demonstrated connexions between their respective assimilation (Kopriva and Rennenberg, 2004; Jamal *et al.*, 2010; Sosa-Valencia *et al.*, 2013). In wheat grain, N supply leads to an increase GSP concentration and duration of accumulation (Pechanek *et al.*, 1997; Triboï *et al.*, 2003; Chope *et al.*, 2014). N and S availability in soil strongly impact GSP composition in mature grain. Indeed GSP differ in their composition in S-amino acids (cysteine and methionine) and S-rich GSP (α/β- and γ-gliadin, LMW-GS) are distinguished from S-poor GSP (ω1,2-, ω5-gliadin and HMW-GS; Shewry *et al.*, 2001). S-deficiency decreases amount of S-amino acids (Wrigley *et al.*, 1980), conducting to a decrease amount of S-rich GSP, while an increase amount was observed for S-poor GSP (Zhao *et al.*, 1999; Wieser *et al.*, 2004; Zörb *et al.*, 2010). N and S are therefore important to control GSP composition and grain N-to-S ratio, generally varying from 12 to 25 (Randall *et al.*, 1981), is a good indicator of grain quality (Wrigley *et al.*, 1980).

But since many years, S deficiency in soils is reported (Tisdale *et al.*, 1986; Zhao *et al.*, 1999). This is explained by the reduction of industrial SO_2 emissions and by a supply of N to the culture without S addition, increasing N-to-S ratio and consequently conducting to S-deficient condition. Besides impacting GSP composition, S deficiency also affects plant growth, decreases yield (Zhao *et al.*, 1999; Hawkesford, 2000) and reduces plant resistance to biotic and abiotic stresses (Rausch and Wachter, 2005; Kruse *et al.*, 2007) It makes S one of the most limiting nutrients for agricultural production (Eriksen, 2009). Since many years, excessive supply of N to the crop in order to increase yield causes environmental perturbations, like water eutrophication (Hirel *et al.*, 2007). To overcome this problem, directives have been implemented to reduce N input to the crop. N and S supply are therefore to be reconsidered and represent an agronomic issue for wheat grain quality.

In this context, it is necessary to identify and understand regulatory mechanisms involved in the grain response to nutrition that influence storage accumulation and particularly GSP accumulation. Because physiological responses rely on complex networks of interactions at different scales, one strategy to meet this objective is to use large-scale 'omics' analysis, and combine information obtained at transcript, protein and metabolite levels (Fukushima *et al.*, 2009). In addition, recent advances in high-throughput technologies and tool development for data integration allow the inference of regulatory networks in plants (Moreno-Risueno *et al.*, 2010; Bassel *et al.*, 2012). Nowadays, this approach is used to study plant responses to stress (Urano *et al.*, 2010), including nutritional stresses (Hirai *et al.*, 2004; Dai *et al.*, 2015). In bread wheat, genes involved in transport and metabolism were identified in the grain response to nutrition, which influenced accumulation of several amino acids, suggested to determine GSP expression (Dai *et al.*, 2015). We assume that post flowering N and S supply disturbs cell metabolism and functioning of the wheat grain at different levels, by modifying several gene expression, protein and metabolite amounts, conducting to changes in GSP composition.

The aim of our study was to obtain an integrative response of the grain to the N and S supply. We used einkorn (*Triticum monococcum*), a wheat diploid species whose genome is close to the A genome of breed wheat (*T. aestivum*, Marcussen *et al.*, 2014). Four treatments containing different N and S amounts were applied to an einkorn culture after flowering state and grains were collected at different times after nutritional change to perform transcriptome and metabolome analyses. Networks generated from the 386 differentially expressed genes, 14 metabolites significantly impacted by nutrition, supplemented by proteomic data previously obtained (Bonnot *et al.*, submitted), have revealed co-expression/accumulation of gene products involved in storage, amino acid metabolism, sulfur metabolism and transcription processes. S deficiency highly perturbed gene expression and protein accumulation. Despite connections between N and S assimilations, few gene products were similarly impacted by these two elements (i.e. increased by N and S or decreased by N and S). Several genes involved in transcription regulation could be central to the network response and could influence GSP accumulation.

RESULTS

Transcriptome analysis revealed 71,5% of expressed genes in the einkorn grain

To investigate effects of post flowering N and S supply on the einkorn grain at transcriptional, metabolic and protein level, four treatments were applied to a *T. monococcum* culture: a low N and S treatment called N-S-, a high N treatment called N+S-, a high S treatment called N-S+ and a fourth treatment with high N and S amounts, called N+S+. Previous results obtained at protein level revealed that N and S modified GSP composition in mature grain, explained by changes in the rate and/or the duration of their accumulation (Bonnot *et al.*, submitted). These results were associated to changes observed for several nuclear proteins and albumin-globulin including some appearing as central actors to the proteome response.

The biological material used in Bonnot *et al.* (submitted) was used for a transcriptome analysis based on RNA sequencing from 54 samples corresponding to six thermal time after anthesis (100, 200, 300, 400, 500, 600°Cd) four treatments (N-S-, N+S-, N-S+, N+S+) from 300 to 600°Cd after anthesis, and three biological replicates. Paired-end sequencing produced 16,386,593 reads per sample in mean. Reads were then mapped against the *T. monococcum* genome at ftp://ftpmips.helmholtz-muenchen.de/plants/wheat/IWGSC/IWGSC_genePredictions_of_other_wheat_species/, which contains 32,047 structurally annotated genes. In mean, 94.2% of the reads were successfully mapped at unique mapping position. Then only genes with Counts Per Million (CPM) value > 1 were considered as expressed (Table S1). A total of 22,910 (71.5%) genes was expressed in at least one sample. From these genes, a large majority was expressed in all conditions and 9,265 (40%) were expressed in all samples (Figure 1a). However, the number of expressed genes was somewhat different between stages and/or nutrition treatments (Figure 1b). For instance, the number of expressed genes at 600°Cd after anthesis was lower for all treatments than in other stages.

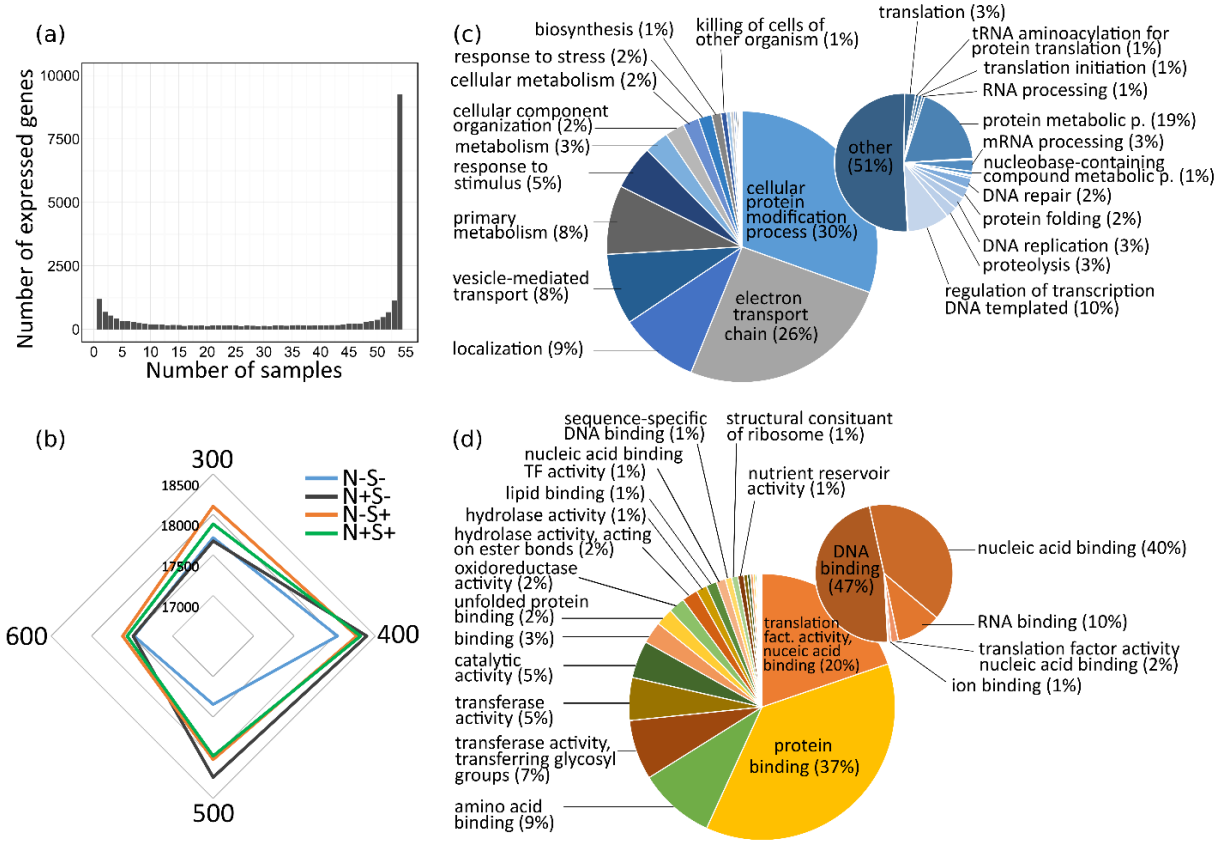


Figure 1. Number of expressed genes in einkorn grain and their functional classification. (a) Histogram showing number of expressed genes per number of samples. (b) Radar plot of number of expressed genes per thermal time after anthesis (300, 400, 500 and 600°Cd) and per nutrition (N-S-, N+S-, N-S+, N+S+). Expressed genes are classified into biological processes (c) and molecular functions (d) according to their Gene Ontology annotation.

Gene ontology was searched for the 22,910 expressed genes. They are classified into 25 biological processes and 31 molecular functions (Figure 1c and Figure 1d, respectively).

For biological processes, the biggest class corresponded to cellular protein modification process (30%) that includes genes involved in protein metabolic process, regulation of transcription, proteolysis, mRNA processing and others. The second biggest process was electron transport chain process (26%).

Then many genes were classified in localization (9%), primary metabolism (8%), vesicle-mediated transport (8%) and response to stimulus (5%). Protein binding was the molecular function most represented among the expressed genes (37%), followed by translation factor activity, nucleic acid binding (20%), which mainly corresponds to genes involved in DNA or nucleic acid binding. Then many genes possess amino acid binding (9%) and transferase (12%) activities.

N and S significantly impacted 386 genes during grain filling

N and S supply highly impacts GSP accumulation during grain filling. So we expected changes at transcript level for genes involved in amino acid metabolism and storage accumulation. From the 22,910 expressed genes, 386 (1.7%) were differentially expressed between at least two treatments and at one time point during grain filling (from 300 to 600°Cd after anthesis, Tables S1 and S3). Principal component analysis was performed from expression data for these 386 genes (Figure 2a). First axis explained 37.2% of total variation and clearly separates the time points of the grain development (300, 400, 500 and 600°Cd). Second axis explained 22.3% of the variation and separates the N+S- treatment from others, particularly from N-S+ treatment.

Hierarchical clustering analysis based on principal components summarized these different observations (Figure 2b). Three clusters were built. Clusters 1 and 3 grouped 182 and 109 genes respectively, whose expression level was strongly determined by developmental stages. Cluster 1 included genes whose expression decreased during grain development while opposite profile was found for genes in cluster 3. Some differences between treatments were found in these two clusters. For example, the cluster 3 included genes expressed, in average, more intensely in the N+S+ treatment than in the N+S- treatment. Cluster 2 grouped 95 genes whose expression level was clearly increased with the N+S- treatment, from 400 to 600°Cd after anthesis. The biggest difference is observed for the comparison between N+S- and N-S+ treatments.

Gene ontology enrichment was then investigated for the 386 genes significantly impacted by the nutrition, compared to all expressed genes (Table 1). Thereby, significant genes were enriched in seven biological processes, corresponding to plant defense (defense response, killing of cells of other organism, heat acclimation), metabolism (cellular catabolic process, asparagine biosynthetic process, lipid phosphorylation) and transport (organic acid transport, Table 1) processes. Enrichment was also observed for 14 molecular functions related to storages and degradation inhibition (nutrient reservoir activity, peptidase inhibitor activity, serine-type endopeptidase inhibitor activity), and binding (zinc ion binding, phospholipid binding, lipid binding, chitin binding, Table 1).

N and S impacted metabolite pools during einkorn grain development

To investigate effects of N and S on amino acid pools and carbon (C) and S metabolisms, which could influence storage accumulation in grain, several metabolites were quantified. We measured from 100 to 950°Cd after anthesis 15 amino acids, five sugars, five organic acids, the S-containing tripeptide glutathione, amine derivatives and nucleosides or derivatives (Table S2). Statistical analysis revealed that 14 metabolites were impacted by nutrition. They were replaced in their respective pathway and their variations observed during grain development and with the four treatments are schematically represented in Figure 3.

N, applied alone or with S (N+S- and N+S+ treatments) have clearly increased total amino acid pool in the grain, compared to the low N treatments (N-S- and N-S+). In details, major effects were observed for aspartic acid (Asp), lysine (Lys) and glutamine (Gln) that showed fairly similar response profiles as they were both increased with N+S- and N+S+ treatments (Figures 3 and S1).

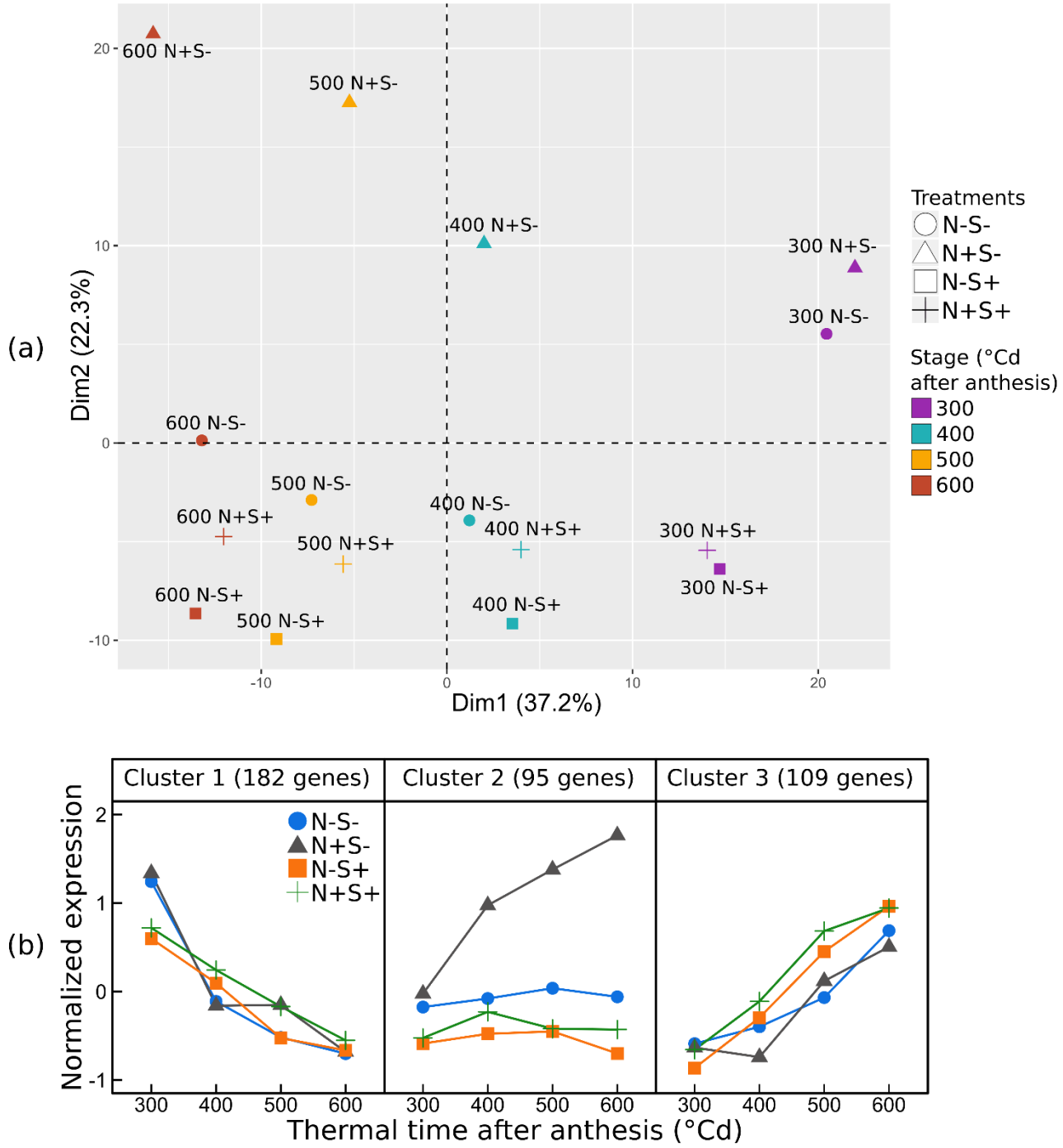


Figure 2. Clustering analysis of differentially expressed genes in response to N and S supply. Principal component analysis (a) from expression data of differential genes allowed the establishment of three expression profiles (b). The number of genes in each cluster is indicated at the top of each panel. Line represent the mean for each nutrition.

Table 1: GO enrichment analysis of Biological process and Molecular function terms for the 386 differential expressed genes in response to nutrition. GO terms enriched in the set of genes that were significantly impacted by nutrition are indicated. For each term, number of genes present in all expressed genes and in significant genes lists are indicated in 'Annotated' and 'Significant' columns, respectively. Number of genes expected in the list of significant genes is indicated in 'Expected' column. Fisher's exact tests were performed to investigate significant difference in GO term enrichment between annotated and significant genes, and *P*value are indicated.

Biological processes					
GO ID	Term	Annotated	Significant	Expected	*P* value
GO:0031640	killing of cells of other organism	10	4	0.19	2.3e-05
GO:0044248	cellular catabolic process	117	9	2.20	0.00031
GO:0006952	defense response	102	8	1.92	0.00060
GO:0006529	asparagine biosynthetic process	4	2	0.08	0.00205
GO:0015849	organic acid transport	49	4	0.92	0.01304
GO:0046834	lipid phosphorylation	10	2	0.19	0.01428
GO:0010286	heat acclimation	1	1	0.02	0.01881
Molecular functions					
GO ID	Term	Annotated	Significant	Expected	*P* value
GO:0045735	nutrient reservoir activity	36	9	0.59	4.4e-09
GO:0008061	chitin binding	2	2	0.03	0.00027
GO:0030414	peptidase inhibitor activity	23	4	0.38	0.00049
GO:0046912	transferase activity, transferring acyl groups	3	2	0.05	0.00080
GO:0008289	lipid binding	97	7	1.60	0.00106
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	224	11	3.70	0.00115
GO:0004867	serine-type endopeptidase inhibitor activity	5	2	0.08	0.00262
GO:0005543	phospholipid binding	31	3	0.51	0.01409
GO:0004134	4-alpha-glucanotransferase activity	1	1	0.02	0.01650
GO:0004781	sulfate adenylyltransferase (ATP) activity	1	1	0.02	0.01650
GO:0008270	zinc ion binding	432	13	7.13	0.02567
GO:0003852	2-isopropylmalate synthase activity	2	1	0.03	0.03274
GO:0046524	sucrose-phosphate synthase activity	2	1	0.03	0.03274
GO:0005274	allantoin uptake transmembrane transporter activity	3	1	0.05	0.04870

Glutamic acid (Glu) and asparagine (Asn) were more abundant in N+S+ condition during grain filling, with an early effect on Glu from 400°Cd after anthesis. Surprisingly, treatments containing high level of N, S or both negatively impacted isoleucine (Ile) and tryptophane (Trp) amounts during grain development. Ile was impacted early in the development while Trp was impacted later, mostly during grain maturation. S-deficiency caused in the N+S- condition decreased S-amino acid methionine, S-tripeptide glutathione and malic acid, an organic acid that is part of the TCA cycle, necessary in the synthesis of the amino acid precursor 2-oxoglutarate (Araujo *et al.*, 2014).

N+S- condition also decreased saccharose amount in grain. Other quantified sugars glucose, fructose, raffinose and trehalose were not affected by treatments, as previously reported on wheat grain (Dai *et al.*, 2015). Trigonelline is an amine derivative and is associated to several physiological roles in plants, including response to oxidative stress, osmoregulator role or could represent NAD reserve during grain germination (Shimizu and Mazzafera, 2000). Interestingly, trigonelline was strongly decreased at grain maturity in all treatments, except with N+S+. In this condition, a very high level of trigonelline was observed especially from 700°Cd to maturity.

Co-expression network reveal highly interconnected genes, metabolites and proteins

N and S supply have modified expression level for 386 genes and the amount of 14 metabolites. Previous results that we obtained from the same biological material have underlined the impact of nutrition for 209 proteins including GSP, nuclear proteins and albumins-globulins. All the -omic data obtained allowed us to investigate relations between transcripts, metabolites and proteins during filling, using network analysis.

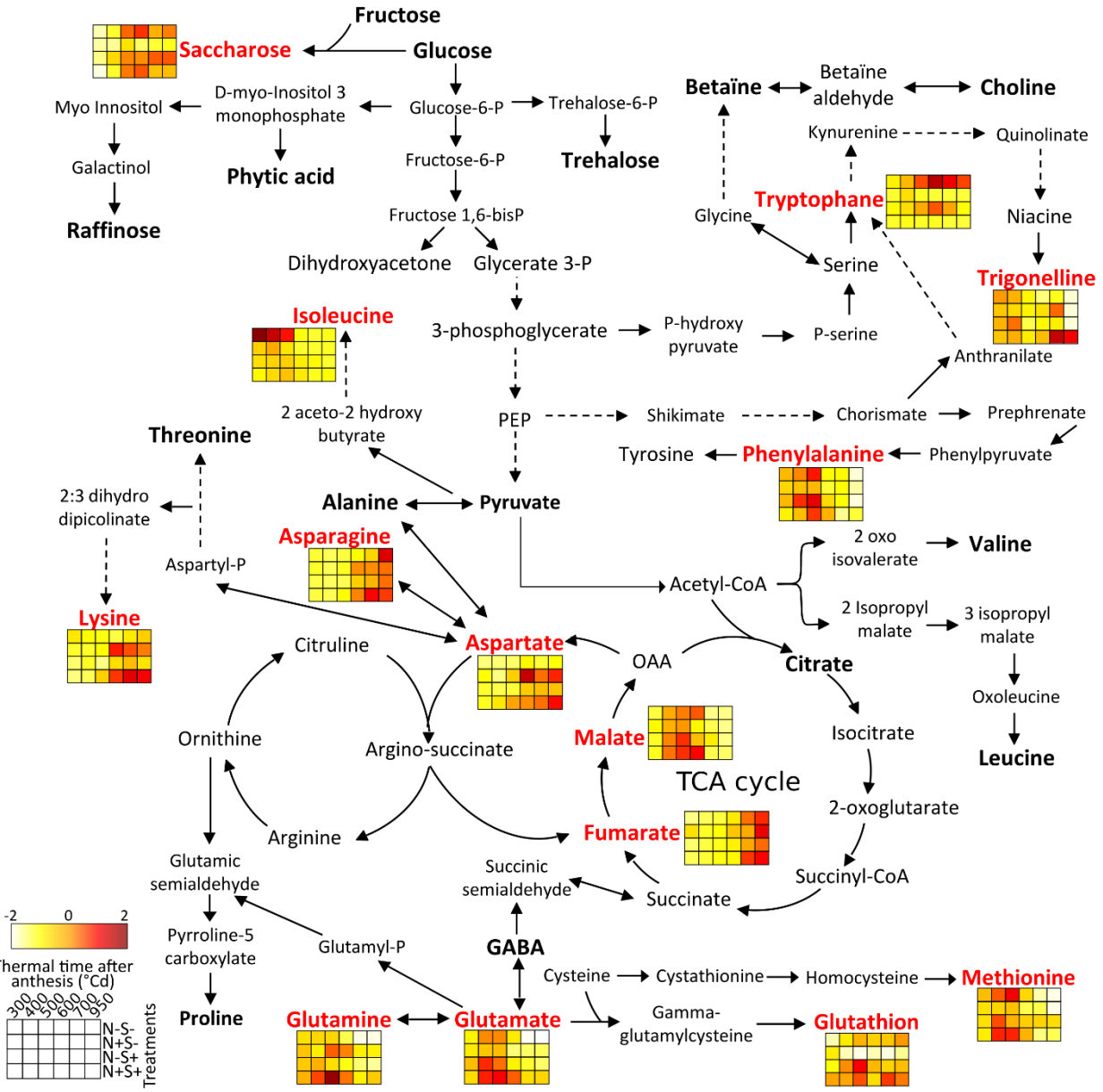


Figure 3. Effects of N and S on several metabolite amounts during einkorn grain development. Amino acids, sugars and organic acids written in bold were those quantified and metabolites with significant nutrition effect are indicated in bold and red. Each is represented in its corresponding pathway. Mean of scaled mass per grain of the 14 significant metabolites are represented on a 4 x 6 grids, with rows and columns representing the four N and S treatments and the six thermal time after anthesis, respectively. The color of each plot represents the mean of the scaled mass of a given metabolite according to a color scale from yellow (low scaled mass) to red (high scaled mass).

First, co-expression/co-accumulation between the different variables was studied. As described in Bonnot *et al.* (submitted), a directed multilevel network, based on the discovery of association rules between attributes, was constructed using all the -omic data significantly impacted by the treatments (i.e. 386 transcripts, 209 proteins, 14 metabolites), the grain N and S contents (GNC and GSC, respectively) and grain dry weight (DW) variables. Associations between these variables were built according to the amount of each variable in the different samples.

The resulting network comprised 189 variables (nodes) and 1634 linkages (edges), distributed into five modules (Figure 4, Table S3). The biggest module (Figure 4a) was composed of 97 variables highly interconnected since they were involved in 1114 linkages (68%). They mainly corresponded to storage compounds attributes, with all GSP (LMW-GS and HMW-GS, ω1,2-, ω5- and γ-gliadin) except α/β-gliadin, several genes and proteins involved in carbohydrate metabolic process and other with defence role and degradation inhibition function. It also contained GNC, GSC and DW. Interestingly and only in this module, three groups can be distinguished based on their linkage number: a core constituted of 20 variables, mainly proteins, with high linkage number (44 to 64 edges each), a second group with 37 variables including four GSP and N and S content, with an intermediate linkage number (21 to 39 edges each) and a third group constituted of 41 variables with low linkage number (1 to 12 edges each, Figure S2a). Two other small modules containing genes mainly involved in similar functions than in module (a) (nutrient reservoir activity, defence, degradation inhibition) were constituted (Figure 4b and c).

Module (d) contained 22 nodes, with several genes involved in amino acid and S metabolisms, with two genes of glutathione S-transferase, a taurine catabolism dioxygenase, a homocysteine S-methyltransferase, a serine hydroxymethyltransferase and an aminotransferase (Figure 4d). Interestingly, central to this module was found a membrane transporter EamA-like (G51, Table S3) that was linked to 14 genes and proteins.

The second largest module comprised 53 variables and 464 edges (Figure 4e). This module (e) is very different from module (a) regarding functions. It comprised many genes or proteins involved in DNA metabolism, transcription, RNA processing or protein modification. Indeed, in this module (e) three DNA replication genes, two RNA helicases, a DNA polymerase, several genes or proteins with DNA or nucleic acid binding activities, five kinases/phosphatase were found, among others. Four proteins and a gene involved in carbohydrate metabolic process and the saccharose were also part of this module (e).

N and S not concomitantly applied highly perturbed grain cell functioning during filling

To group genes, metabolites and proteins strongly and similarly impacted by nutrition during filling phase, another network was built (Figure 5). Transcript level or metabolite/protein amount measured in the different samples from 300 to 500°Cd after anthesis were compared in the different treatment conditions. Attributes more expressed/accumulated in one treatment compared to another were linked to the corresponding treatment comparison. The corresponding network comprised the most significant variables in at least one treatment comparison. It included 499 linkages (edges) and 275 variables (nodes) including four GSP, seven metabolites, 178 transcripts and 86 proteins (70 albumin-globulin and 16 nuclear proteins).

First, the most striking observation was the disparity in the linkage number for the different treatment comparisons (Figures 5 and S2b). Indeed, the three comparisons containing the N+S- treatment were those with high linkage number, suggesting that this treatment highly differ from others regarding its effect on the einkorn grain. 165 variables showed significant difference between the N+S- and N-S+ treatments, highlighting that these two conditions had the most different effects. Interestingly, as previously observed at protein level, very few attributes (44) were linked to the N-S-/N+S+ comparison. So these two treatments did not have very different effects, whereas they had the most contrasting N and S amounts. Similar results were observed for N-S+/N+S+ and N-S-/N-S+ comparisons, each linked to 33 and 22 variables, respectively.

Many attribute groups were constituted on this network, which showed similar response to the nutrition. Approximately half of attributes (137) was linked to only one treatment comparison, showing that their abundance only differs between two nutrition conditions. Other attributes were associated to two or more treatment comparisons, indicating the strong effect of N and S supply on these variables, which showed very different quantities in the four nutrition conditions.

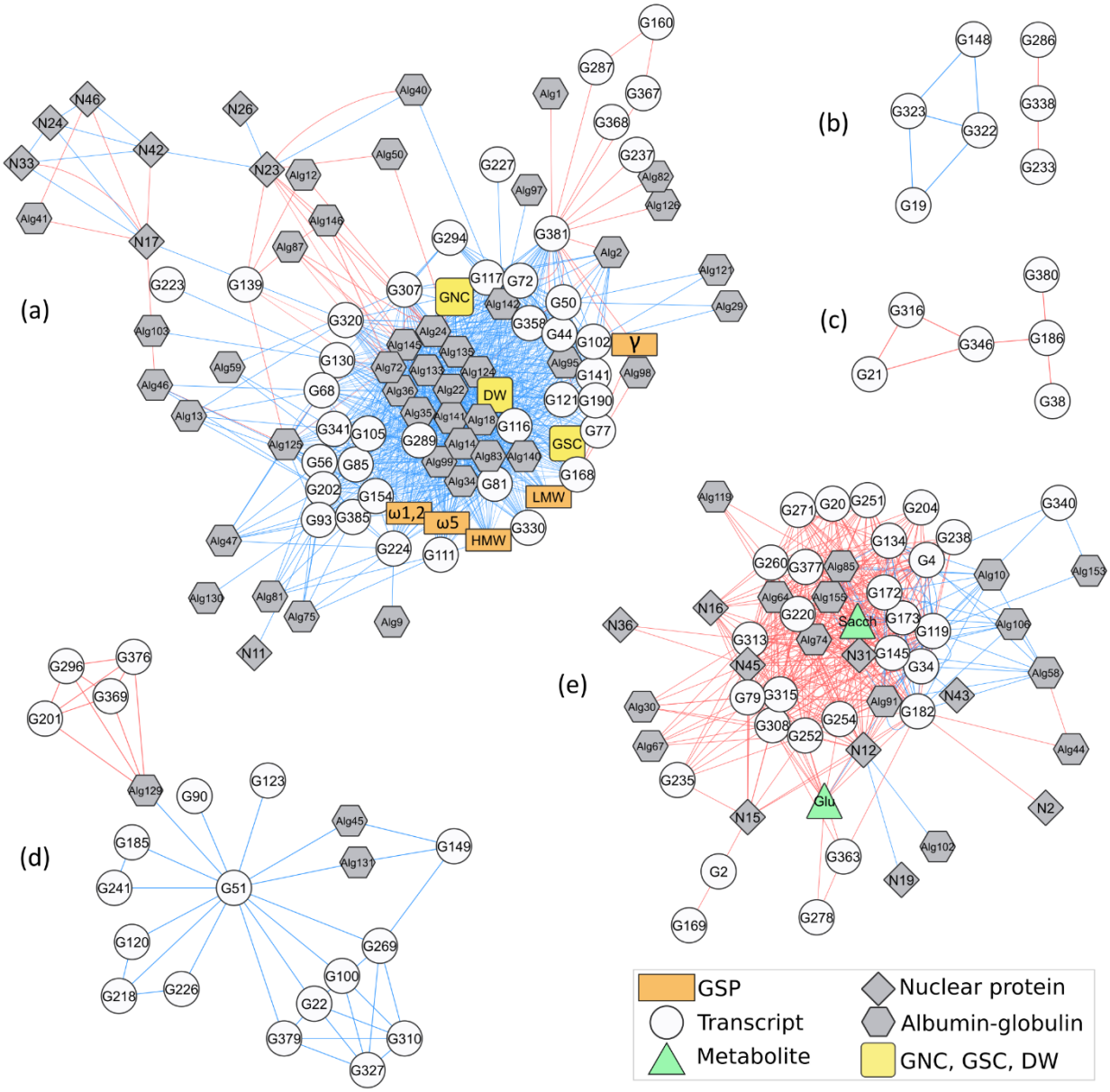


Figure 4. Networks of co-expression/accumulation during grain development of transcripts, metabolites and proteins impacted by N and S supply. Node shape and color reflect attribute category: white circles, transcripts; green triangles, metabolites; orange rectangles, GSP; grey diamonds, nuclear proteins; grey hexagons, albumins-globulins; yellow squares, N and S content (GNC and GSC, respectively) and dry weight (DW). Edge colour indicates the biological significance between source and target, red when the quantity of RNA/metabolite/protein was high for the two nodes and blue when the quantity was low for the two nodes. Codes were attributed for genes and proteins, for details see Table S1. Sacch: saccharose; Glu: glutamate; Asp: aspartic acid.

In this case, one group stood out, composed of 47 attributes including HMW-GS and N-to-S ratio variables which were all increased by N+S- compared to the three other treatments. Interestingly, 16 attributes in this group were part of the 22 variables that composed module (d) in co-expression network (Figure 4d, Table S3). The corresponding genes and proteins, mainly involved in amino acid and S metabolism were thus strongly increased by high level of N and low level of S. In addition, HMW-GS showed a strong and early increased rate of accumulation in the N+S- condition. In this group of 47 attributes, two sulfate transporters were also found, as well as an amino acid transporter and two cysteine synthase, highlighting the activation of S metabolism and amino acid metabolism/transport in S-deficiency condition caused with the N+S- treatment.

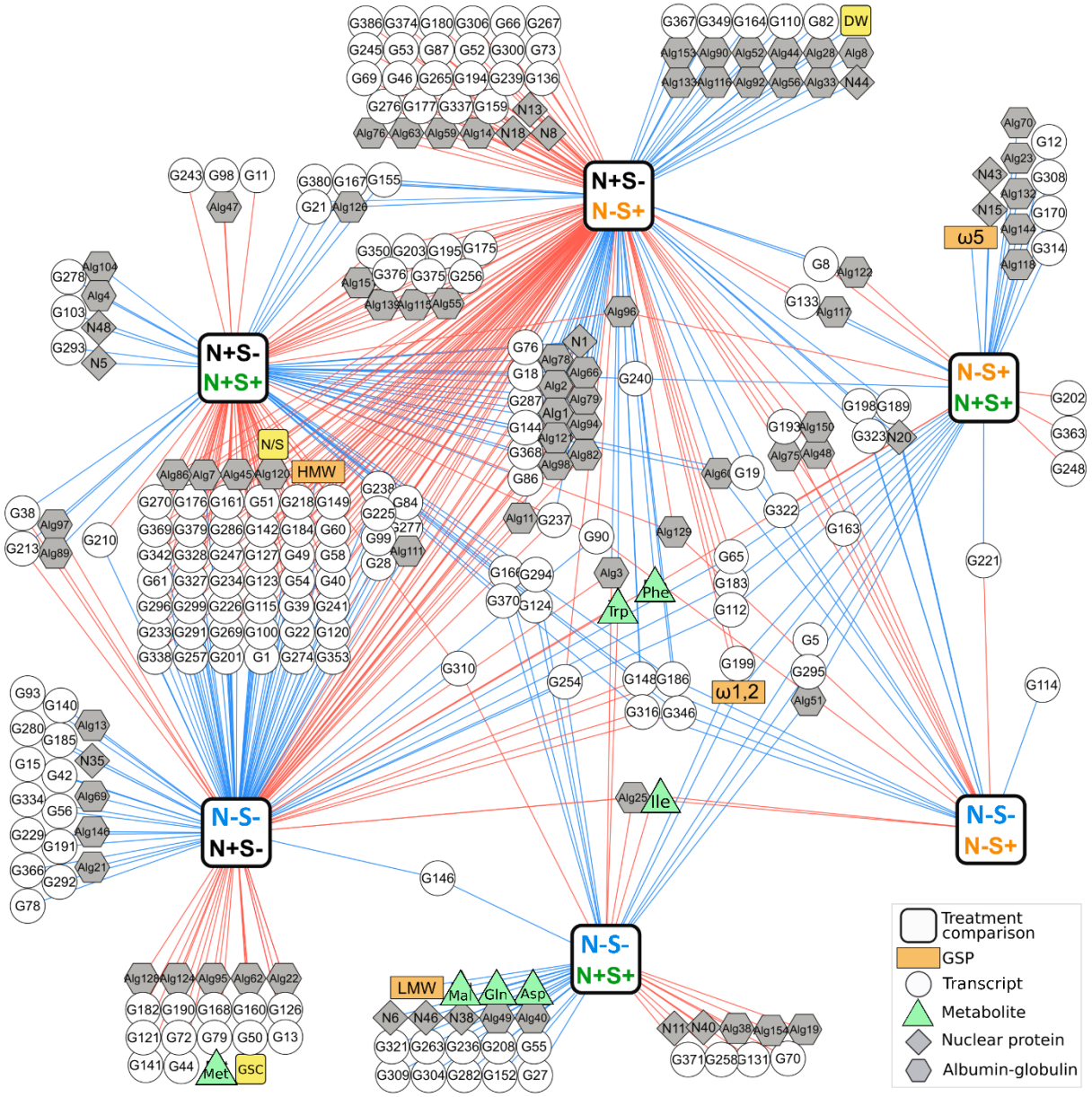


Figure 5. Network analysis for gene, metabolite and protein response to N and S nutrition. Genes differentially expressed and metabolites and proteins differentially accumulated in response to the nutrition were related to the six treatment comparisons. Node shape and color reflect attribute category: black squares, treatment comparisons; white circles, genes; green triangles, metabolites; yellow rectangles, GSP; grey diamonds, nuclear proteins; grey hexagons, albumins-globulins, orange squares, N and S content (GNC and GSC, respectively) and dry weight (DW). Edge colour indicates the biological significance between source and target, red when the quantity of transcript/metabolite/protein was higher in the first treatment used in the comparison and blue when the quantity was lower in the first treatment. Codes were attributed for genes and proteins, for details see Table S1. Phe: phenylalanine; Trp: tryptophane; Ile: isoleucine; Met: methionine; Mal: malic acid; Gln: glutamine; Asp: aspartic acid.

The two generated networks (co-expression and nutrition networks) allowed us to point that 46% of co-expressed attributes (Figure 4) were also found in the nutrition network (Figure 5). The module (d) in co-expression network is present at 95% in the nutrition network. From these 22 variables, 16 (72%) were in the same group, which was strongly increased in the N+S- condition.

Then the combined effects of N and S supply on the einkorn grain were represented in a simplified scheme (Figure 6). Based on network analysis, it summarized the antagonist effect of N and S evidenced in present study. Indeed, it was found that few variables were increased or decreased both with N and S. However, this was the case for several genes involved in transport process, oxido-reduction, transcription and for Asp, Mal and the two GSP, LMW-GS and α/β-gliadin. In the same way, Ile and genes involved in response to stress, carbohydrate metabolic process and translation were biological processes both decreased by N and S. But a large majority of variables showed very different responses to N or S. From this observation it was shown that many genes and proteins acting in S and amino acid metabolism (including module [d]), transport and transcription were increased with N and decreased with S. Conversely, among those increased with S and decreased with N, nutrient reservoir activity, carbohydrate metabolic process and response to stress are processes that were highly represented.

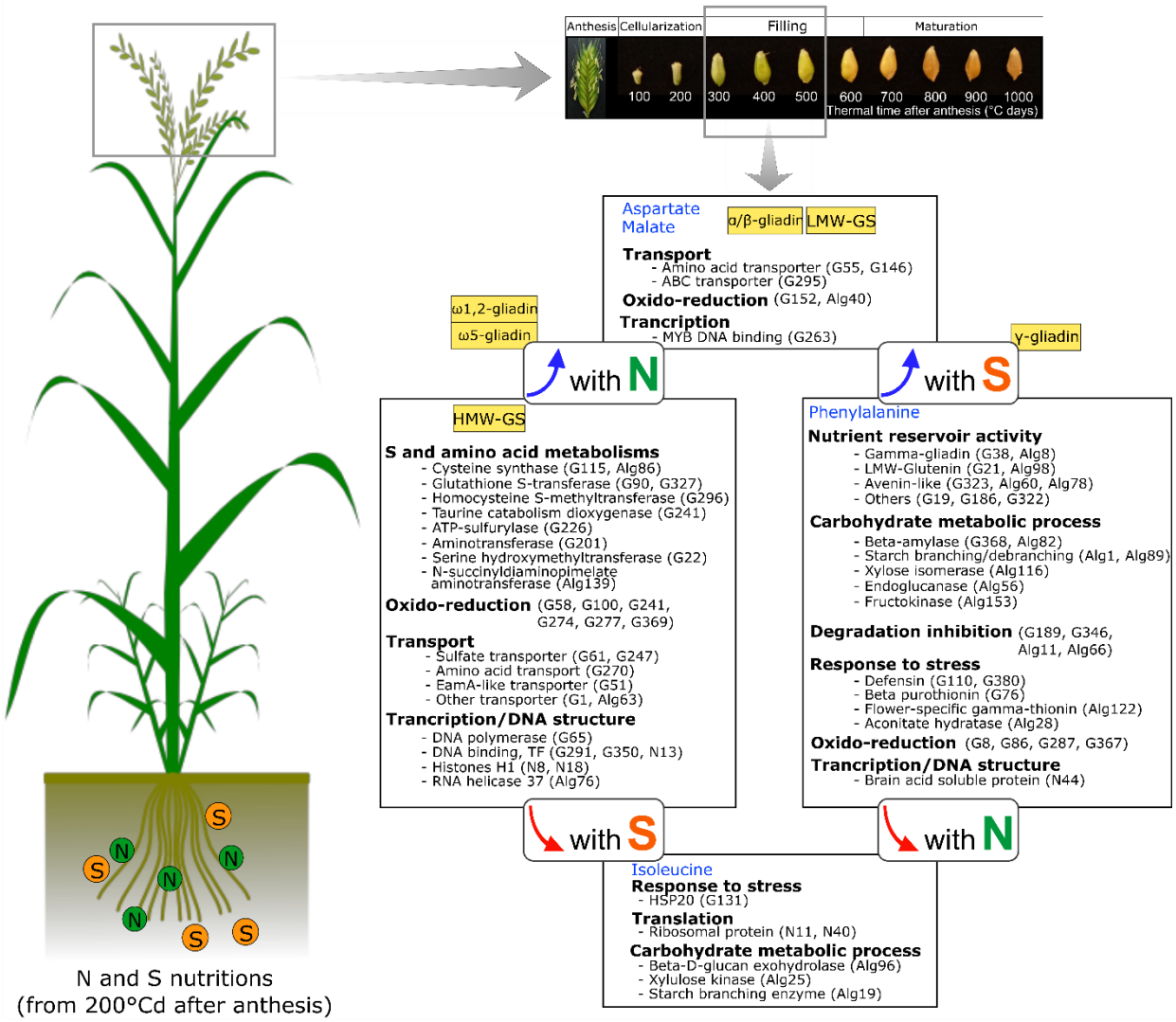


Figure 6. Overview of main processes impacted by N and S supply during filling. Genes, metabolites and proteins presented are part of the network present in Figure 5.

DISCUSSION

In wheat, the GSP composition defines dough properties and thus the baking quality. Identification of genes and molecular mechanisms impacted by N and S supply in the wheat grain appears necessary to understand the importance of nutrition, known for a long time for its effect on the GSP composition in mature grain. The aim of present study was to obtain an integrative view of the grain response to N and S supply and highlight genes and mechanisms that could be essential in this regulation. To this end, we have conducted an experiment in einkorn. Analysis of metabolites and transcriptome, combined to results previously obtained from the same biological material at protein level, allowed us to show the high impact of N supply which increased amino acid metabolism and also S metabolism and transport, probably to compensate for low S availability. We therefore observed that major changes in gene expression or metabolite/protein accumulation were explained by a non-concomitant supply of N and S to the culture, which highly perturbed the N-to-S ratio.

High effect of N nutrition highlighted the strong incidence of S-deficiency on the einkorn grain

In present study, the four treatments applied to the einkorn culture corresponded to a low N and S condition (N-S-), a high N treatment (N+S-), a S treatment (N-S+) and a fourth treatment with a concomitant supply of N and S (N+S+). These different conditions have clearly modified GSP composition at maturity, explained by an early modification of rate of accumulation during filling (Bonnot *et al.*, submitted). Since GSP differ in their composition in S-amino acids, N-to-S ratio in grain is responsible for the balance between S-rich and S-poor GSP synthesis. In our experiment, N-to-S ratio was strongly increased in the N+S- condition, from 300°Cd after anthesis and it exceeded 30 at 600°Cd. This corresponded to a S-deficient condition due to a high amount of N without concomitant S supply (Zhao *et al.*, 1999; Eriksen, 2009). This condition leads to decrease gliadin-to-glutenin ratio, known to be negatively correlated to bread loaf volume (Dhaka and Khatkar, 2015). Moreover, a decreased amount was also observed for the S-tripeptide glutathione, known to modify the GSP polymerization and thus impact dough properties (Tea *et al.*, 2005).

More generally, N+S- treatment was shown to be very different from others in the different analysis. Principal component analysis and hierarchical clustering performed on the 386 significant genes highlighted in the transcriptome analysis revealed a clear distinction of this condition from others. This observation was confirmed with the results obtained from all -omic datasets. In network analysis, many genes and proteins and several metabolites were connected to the treatment comparisons which included N+S- condition. Among them, some were strongly increased in this condition, compared to the three others. Several were part of the same module, which showed their strong co-expression/accumulation. Their functional annotation supposed an activation of amino acid metabolism and transport with a high level of N. This is in accordance with results obtained at metabolite level, with the increased amount of total amino acid pool in the high N treatments, which was previously observed in several studies (Howarth *et al.*, 2008; Dai *et al.*, 2015). HMW-GS were also strongly and early increased in this high N condition, without any S supply. They are part of S-poor GSP, since S-amino acid represent 0.8% of its amino acid total, while they represent, for example, 4.6% in the S-rich γ-gliadin (Shewry *et al.*, 2001).

Several genes involved in S metabolism and transport were co-expressed and increased in the N+S- condition. This suggests an activation of S assimilation, probably to compensate for S deficiency. This plant response to S-starvation is relatively well-understood and reported in several studies (Nikiforova *et al.*, 2003; Hawkesford and DeKok, 2006).

In our study, S supply was associated to an increased amount of GSP, rate of accumulation of S-rich GSP and generally to an increased expression level and protein accumulation for genes involved in reserve establishment and carbohydrate metabolic process. Taking these results together, we hypothesized that lack of S in a condition of high N supply probably highly perturbed einkorn grain to limit the impact on the amount of reserves available in grain that are essential to the germination process.

Several genes highly perturbed by the N to S ratio modification could be central in controlling GSP composition

In our study, results obtained in the N-S- condition were quite surprising and interesting. Indeed, the lack of N and S supply with this treatment resulted in correct amount and composition of GSP in mature grain. Even though total GSP amount was increased in the N+S+ condition, amount obtained with N-S- was not lower than in the N+S- condition and GSP composition was comparable between N-S- and N+S+ treatments. It is widely understood that the major source of grain N is accumulated before anthesis. Indeed from 50 to 95% of grain N comes from its remobilization from shoots and roots where it was stored (Palta and Fillery, 1995; Kichey *et al.*, 2007). We therefore hypothesized that in our experiment and in the N-S- condition, a sufficient accumulation of N and S before flowering and an activation of their remobilization from the vegetative organs to the grains during development occurred.

In our results, N-to-S ratio was almost the same between N+S+ and N-S- conditions. Many genes and proteins seemed to be increased by N and decreased by S, or conversely, demonstrating an antagonist effect of N and S on these genes. This explains why major differences were observed when only N or only S were applied to the culture. We therefore hypothesize that these genes were regulated by the N-to-S ratio. In addition, network analysis revealed that N-S- and N+S+ did not have very different effect on transcripts, metabolites and proteins, since only 44 variables showed significant changes between these two conditions. Taken together, these results showed that a high N and S supply slightly affected the einkorn grain, while the modification in N-to-S ratio seemed to impact an important number of genes and proteins.

Among them, some could be involved in the plant response and adaptation to a S or N deficiency, to control GSP synthesis and composition. As we mentioned above, activation of S and amino acid metabolisms was observed in S-deficient condition, when N-to-S ratio was clearly increased. In network analysis, co-expression of genes and proteins involved in this response have highlighted a membrane EamA-like transporter, central to the corresponding module (Figure 4d, Table S3). EamA family is not well functionally characterized in plant, but it is part of the drug/metabolite transporter superfamily (Västermark *et al.*, 2011) involved in metabolite efflux, and particularly O-acetylserine and cysteine in *Escherichia coli* (Franke *et al.*, 2003). Thus this transporter probably was central to the grain response and adaptation to S-deficiency and probably acts with the two sulfate transporters which showed a similar response to nutrition.

Several genes with DNA binding activity and involved in transcription process were revealed in present study, including some which showed very different expression level or protein accumulation between N+S- and N-S+ conditions. A CBF transcription factor was more expressed with high level of N than S. CBF transcription factors can act in multiple stress responses as previously reported (Akhtar *et al.*, 2012; Kidokoro *et al.*, 2015). Another gene with similar response profile possessed a plant homeodomain (PHD), thought to be involved in chromatin-mediated transcriptional regulation (Sanchez and Zhou, 2011). These results suggest a control mechanism, triggered by the disruption of the N-to-S ratio, involving transcriptional regulators and transporters for regulating the balance and therefore the grain composition.

Genes, proteins and metabolites which did not emerge from network analysis

In this study, we performed two types of network analysis to investigate co-expression and co-accumulation of genes, proteins and metabolites and to group those similarly impacted by N and S supply. Several variables were both revealed in these two analysis, like attributes found in module (d) and discussed below. But several variables were absent from the co-expression network or the nutrition network or both.

In module (e), 53 highly interconnected variables mainly involved in DNA transcription, replication, mRNA processing, protein modification, among other functions, were revealed. Only 10 (19%) were present in the nutrition network, which contained attributes presenting a clear different expression level or amount between at least two treatments. Thus some variables significantly impacted by nutrition could be absent from network analyses because of the thresholds that we

used for the parameters that define rule quality. However, these variables have to be considered too. Indeed, some transcripts, proteins or metabolites could be slightly but significantly impacted by the nutrition, at one time point during grain development, and yet could have a high impact in the cell. It was previously reported the dynamics of co-expression network across plant development, in different cell types or in response to a nutritional stress (Brady *et al.*, 2007; Wei *et al.*, 2013; Serin *et al.*, 2016). Some genes present in module (e) could correspond to key transcriptional regulators temporarily impacted by N and S, which could conduct to major effect on the grain adaptation to nutritional changes. For example, a gene with OB-fold nucleic acid binding domain was revealed and was connected to 26 other variables.

An ABC transporter was linked to 24 other genes or proteins in module (a) but was absent from the nutrition network. ABC transporters represent one of the largest protein families and are involved in such diverse processes as plant nutrition, response to abiotic stimulus or plant pathogen defence (Kang *et al.*, 2011). Since this ABC transporter was present in module (a), it was co-expressed with several other genes involved in storage accumulation and it therefore could be responsible for adjustment in storage accumulation in response to N and S supply.

Moreover, genes presenting atypical profile, not present in co-expression network can act alone and yet have an important role in the grain cell. In this case, a multiprotein bridging factor (MBF) was impacted by nutrition but absent from networks. Nevertheless MBF can be central in the grain response since its role in response to abiotic stress was previously demonstrated in cereals (Qin *et al.*, 2014). Other transcription factors like a bZIP and MYB-related protein have to be considered too. They all could be involved in a transcriptional reprogramming occurring after N and S supply and modifying storage accumulation in einkorn grain during filling phase.

In present study, our results have evidenced that a non-concomitant N and S supply, which highly disturbed N-to-S ratio was responsible for major changes observed at transcript, metabolite and protein levels. Without forgetting genes slightly impacted by the nutrition, data integration with network analysis allowed us to point several genes central to the grain response, which deserve further investigation.

EXPERIMENTAL PROCEDURES

Plant material

Triticum monococcum ssp. *monococcum* (accession ERGE 35821) seeds were germinated (5 February) at room temperature on filter paper in Petri dishes moistened with demineralized water. When the radicles were 0.5 cm to 1 cm long, seedlings were transplanted to 50 mL PVC columns (7.5 cm i.d. × 50 cm deep; 2 seedlings per column) filled with a 2:1 (v/v) mixture of washed perlite:river sand. The columns were arranged in a greenhouse to form a homogenous stand with a population density of 512 plant m^{-2}. The experimental design was a randomized complete block design with four blocks and four treatments. The high plant density inhibited the development of axillary tillers which coordinated the development of the plants within and between the containers. In the greenhouse air temperature at the top of the plant stand was controlled at 20°C/15°C (day/night) and air relative humidity was controlled at 55%/75% (day/night). Plant received a mean total daily photosynthetic photon flux (PPF) of 152 and 161 µmol m^{-2} d^{-1} between transplanting and anthesis and between anthesis and grain ripeness maturity, respectively, over the light period (16-h light/8-h dark). Air temperature and relative humidity were measured in miniature forced-draft flues placed in each block in the center of the plant stands at the height of the ears. Air temperature, relative humidity, and PPF were measured every min and 15-min averages were recorded on a data logger. Thermal time was calculated by summing degree-days (base temperature 0°C) calculated as average daily air temperature.

Until anthesis, each PVC column received 167 mL d^{-1} of a modified Hoagland's nutrient solution (Castle and Randall, 1987) prepared with demineralized water and containing 3 mM of N and 0.1 mM of S (1 mM KH_2PO_4, 1 mM KNO_3, 0.5 mM $Ca(NO_3)_2$, 0.5 mM NH_4NO_3, 0.1 mM $MgSO_4$, 1.9 mM $MgCl_2$, 3.5 mM $CaCl_2$, 4 mM KCl, 10 µM

H_3BO_3, 0.7 µM $ZnCl_2$, 0.4 µM $CuCl_2$, 4.5 µM $MnCl_2$, 0.22 µM MoO_3, and 50 µM EDFS-Fe). At anthesis, the nutrient solution was replaced by demineralized water to avoid any over accumulation of N and S compounds in the plant or in the potting substrate as during that period N and S plant demand is very limited. Then four combinations of N and S supply were applied from 200°Cd to 700°Cd after anthesis: a low N and S supply treatment corresponding to a nutrient solution with no N and S (treatment termed N-S-); a high N and low S treatment with 6 mM N and no S (N+S-); a low N and high S treatment with 0.5 mM N and 2 mM S (N-S+); and a high N and S treatment with 6 mM N and 2 mM S (N+S+). Nitrogen was not totally excluded from the N-S+ nutrient solution as previous experiments have shown that in the absence of N in the nutrient solution S uptake and/or metabolism are inhibited (P. Martre unpublished data). The nutrient solutions were modified as described in Dai *et al.* (2015). Grains were collected every 100°Cd after anthesis from 100°Cd to full maturity (950°Cd) for metabolite analysis, then from 100 to 600°Cd for transcriptome analysis.

Metabolite assays

Robotized analyses

Metabolites were extracted from about 20 mg of ground fresh weight with ethanol/water and analyzed as previously described (Biais *et al.*, 2014; Bénard *et al.*, 2015). Soluble sugars including sucrose, glucose and fructose (Jelitto *et al.*, 1992), malate (Nunes-Nesi *et al.*, 2007) citrate (Tompkins and Toffaletti, 1982) and the sum of free amino acids were determined in the ethanolic supernatant. Starch content was determined on the pellet as described in (Hendriks *et al.*, 2003). Starch and amino acids contents were determined in 96-well polystyrene microplates (Sarstedt, Marnay, France) using a robotized Starlet platform (Hamilton, Villebon sur Yvette, France). Individual amino acids were analyzed using the AQC-tag method. Oxidized (GSSG) and reduced (GSH) glutathione were extracted and assayed enzymatically with glutathione reductase and 5,50-dithiobis-2-nitrobenzoic acid, as described by Griffith (1980).

^{1}H-NMR analyses

Polar metabolites were extracted from 50 mg of grain lyophilised powder with an ethanol–water series at 80°C (adapted from Moing *et al.* [2004]), Allwood *et al.* [2011]). The supernatants were combined, dried under vacuum and lyophilized. Each lyophilized extract was solubilized in 500 µL of 200 mM deuterated potassium phosphate buffer solution (apparent pH 6.0), containing 3 mM ethylene diamine tetraacetic acid disodium salt (EDTA) to chelate paramagnetic cations and improve spectrum resolution (especially in the citrate region), titrated with KOD 1 M or DCL 0.1 M solutions by means of BTpH (Bruker BioSpin GmbH, Rheinstetten, Germany) to apparent pH 6.00 ± 0.02, and lyophilized again. The lyophilized titrated extracts were stored in darkness under vacuum at room temperature, before ^{1}H-NMR analysis was completed within one week.

For ^{1}H-NMR analysis, 500 µL of D_2O with sodium trimethylsilyl [2,2,3,3-d_4] propionate (TSP, 0.01% mg/mL final concentration for chemical shift calibration) were added to each lyophilized titrated extract. The mixture was centrifuged at 17,700 × *g* for 5 min at room temperature. The supernatant was then transferred into a 5 mm NMR tube for acquisition. Quantitative ^{1}H-NMR spectra were recorded at 500.162 MHz and 300 K on a Avance III spectrometer (Bruker Biospin, Wissembourg, France) using a 5-mm ATMA broadband inverse probe flushed with nitrogen gas and an electronic reference for quantification (Digital ERETIC, Bruker TopSpin 3.0). Sixty-four scans of 32k data points each were acquired with a 90 ° pulse angle, a 6000 Hz spectral width, a 2.73 s acquisition time, and a 20 s recycle delay. The assignments of metabolites in the NMR spectra were made by comparing the proton chemical shifts with literature (Likes *et al.*, 2007) or database values (MeRy-B, HMDB; BMRB, MMCD), by comparison with spectra of authentic compounds and by spiking the samples. For assignment purposes, ^{1}H-^{1}H COSY, ^{1}H-^{13}C HSQC spectra were acquired for a selected sample. For absolute quantification, four calibration curves (glucose and fructose: 2.5 to 100 mM; glutamate and glutamine: 2.5 to 30 mM) were prepared and analysed under the same conditions. The glucose calibration was used for the quantification of all compounds, as a function of the number of protons of selected resonances except fructose, glutamate and glutamine that were quantified using their own calibration curve. The concentration of each

organic or amino acid was expressed as g of the acid form per weight unit. The metabolite concentrations were calculated using AMIX (version 3.9.14, Bruker) and Excel (Microsoft, Redmond, WA, USA) softwares.

RNA extraction and sequencing

RNA were extracted from einkorn grain according to method used by Pont *et al.* (2011). Three biological replicates were used. Grain samples (stages 100, 200, 300, 400, 500, 600°Cd after anthesis, treatments N-S-, N+S-, N-S+, N+S+ for each stage from 300 to 600°Cd after anthesis) were ground in liquid nitrogen and 4.5 mL of buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1 M NaCl, 1% SDS) were added to approximately 0.5g of powder. The supernatant was collected and 3 mL of phenol-chloroform-isoamyl alcohol mixture (25:24:1) were added to eliminate starch and proteins. Supernatant was collected and nucleic acids were precipitated by adding 0.1 vol of 3 M AcNA pH 5.2 and 2 vol of 100% EtOH. Pellet was dried under ambient condition for 1 h and dissolved in RNase-free water. DNAse treatment was then performed using a RNase-Free DNase set (Qiagen) and then a RNeasy MinElute Cleanup kit (Qiagen). The amount and integrity of RNA integrity was checked by agarose gel electrophoresis. Only samples with a RIN greater than 7 were used for the library construction.

RNA Sequencing was done at Eurofins MWG Operon (www.eurofinsgenomics.eu, Ebersberg, Germany). RNA strand-specific libraries were created using commercially available kits according to the manufacturer's instructions. In brief, polyA-RNA was extracted from total RNA using an oligodT-bead based method. After fragmentation of the mRNA, first and dUTP based second strand synthesis was carried out, followed by end-repair, A-tailing, ligation of the indexed Illumina Adapter and digestion of the dUTP-strand. Size selection was done using a bead based method targeting an average insert size of 150-400 base pairs. After PCR amplification the resulting fragments were cleaned up, pooled, quantified and used for cluster generation. For sequencing, pooled libraries were loaded on the cBot (Illumina) and cluster generation was performed using manufacturer's instructions. Paired-end sequencing using 125 bp read length was performed on a HiSeq2500 machine (HiSeq Control Software 2.2.38) using HiSeq Flow Cell v4 and TruSeq SBS Kit v4. For processing of raw data RTA version 1.18.61 and CASAVA 1.8.4 was used to generate FASTQ-files.

Read alignment and expression analysis

Reads were mapped on the *T. monococcum* genome available in MIPS genome database (Mewes, 2002) reduced to scaffolds containing the 32,047 annotated genes and the biggest scaffolds (337,103 scaffolds used). Mapping was performed using BWA-MEM (version 0.7.12-r1039, http://bio-bwa.sourceforge.net/). Only reads with unique mapping position and a mapping quality score of at least 10 were considered for read counting. Paired-end reads that were mapped to the same reference with about the expected insert size were counted as one read. Paired-end reads that were mapped to different references of with an unexpected insert size were counted as two reads. If only one read of a pair was mapped, it was counted as one read. Single-end reads were used straightforwardly. Only reads overlapping exon-features were counted. All reads mapping to features with the same grouping-tag were summed. A Trimmed Mean of M-values (TMM) normalization was performed using the edgeR package (Robinson and Oshlack, 2010; Robinson *et al.*, 2010) to reconstruct transcripts and estimate transcript abundance in counts-per-million (CPM). Genes with CPM value higher than 1 were considered as expressed.

Statistical analysis

For RNA sequencing data, differential expression analysis was performed using the EdgeR package (Robinson and Oshlack, 2010; Robinson *et al.*, 2010). Genes with a CPM superior to 1 in at least three samples were considered. EdgeR estimates the genewise dispersions by conditional maximum likelihood, conditioning on the total count for that gene. An empirical Bayes procedure was used to shrink the dispersions towards a consensus value, effectively borrowing information between genes. At each thermal time after anthesis from 300 to 600°Cd, differentially expressed genes among the four N and S treatments were identified using an exact test analogous to Fisher's exact test, but adapted for over-dispersed data (excerpt from Robinson *et al.*, 2010). Differential expression was considered to be significant at adjusted P value < 0.001 after false discovery rate correction (Benjamini and Hochberg, 1995).

Principal component analysis performed on differential expressed genes was realized with FactomineR (Husson *et al.*, 2015) package for R and hierarchical clustering on principal components was performed with the 'hcpc' function and the 'ward' method to build gene expression profiles. Number of clusters was determined with the suggested partition, corresponding to the one with the higher relative loss of inertia. R package ggplot2 (Wickham, 2011) was used to draw plots.

Metabolite data were analysed using the statistical software R v3.2.2 (R Core Team, 2015). Two-way (developmental stage cross nutrition) ANOVA, followed by Tukey's honestly significantly different (HSD) test were performed. Statistical differences were judged at P value < 0.05.

Functional annotation and gene ontology enrichment analysis

Annotation of differential expressed genes in response to nutrition was performed using TriAnnot (Leroy *et al.*, 2012) pipeline. Gene ontology (GO) enrichment was then performed, as described in Choulet *et al.* (2014). Briefly, GO terms were searched for all expressed genes by BLAST (e-value < 1e-05), performed against the PLAZA 2.5 database (Proost *et al.*, 2009). Only the five best hits with more than 50% coverage were used for the analysis. Enrichment calculations were then performed using the topGO R package (Alexa and Rahnenfuhrer, 2010). GO terms for expressed genes were used as the reference comparison set against differential expressed genes. Fisher's exact test were performed and statistical difference in GO term enrichment was considered at P value < 0.05.

Network analysis

Data set used for network analysis comprised the 386 genes and 14 metabolites which showed significant nutrition effect in present study, supplemented by 6 GSP and 203 proteins previously obtained and variant in response to nutrition from the same biological material (Bonnot *et al.*, in prep). Network were built using RulNet platform ((http://rulnet.isima.fr; Vincent *et al.*, 2015). An algorithm describing rule semantics between attributes (Agier *et al.*, 2007) is used to generate rules among -omic data, and make possible to define different semantics. Data were scaled and a semantic was written which allowed to find rules between co-expressed/accumulated RNAs, metabolites and proteins (First semantic, Method S1). Three quality measures were used to select the best rules (support, confidence and lift) and rules with a support >0.3, confidence = 1 and a lift >2 were selected. A second semantic was used to find rules between RNAs/metabolites/proteins and comparisons of treatments, according to their quantity in the different samples (Second semantic, Method S1). For this second semantic, confidence and lift thresholds were set at 0.7 and 1.5, respectively. Rules validated allowed to build network using CYTOSCAPE software version 3.3.0 (Smoot *et al.*, 2011).

ACKNOWLEDGEMENTS

We thank Richard Blanc (UMR GDEC, INRA, Blaise Pascal University) for help in the greenhouse and David Alvarez, Annie Faye, Marielle Merlino, Sibille Perrochon and Isabelle Nadaud for help collecting grains (UMR GDEC). We thank Florent Murat for his help to adapt the *T. monococcum* reference genome used in the mapping step and Eurofins Genomics for performing RNA sequencing analyses. This work was supported by a Ph.D. grant from the French Ministry for Higher Education and Research to TB and funding from the French Government managed by the National Research Agency (ANR) in the framework of Investments for the Future (ANR-10-BTBR-03), France AgriMer and the French Fund Supporting Plant Breeding (FSOV).

REFERENCES

Agier, M., Petit, J.M. and Suzuki, E. (2007) Unifying framework for rule semantics: Application to gene expression data. *Fundam. Informaticae*, 543–559.

Akhtar, M., Jaiswal, A., Taj, G., Jaiswal, J.P., Qureshi, M.I. and Singh, N.K. (2012) DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. *J. Genet.*, **91**, 385–395.

Alexa, A. and Rahnenfuhrer, J. (2010) topGO: Enrichment analysis for Gene Ontology.

Allwood, J.W., Vos, R.C.H. De, Moing, A., Deborde, C., Erban, A., Kopka, J., Goodacre, R. and Hall, R.D. (2011) *Plant metabolomics and its*

potential for systems biology research: Background concepts, technology, and methodology 1st ed., Elsevier Inc.

Araujo, W.L., Martins, A.O., Fernie, A.R. and Tohge, T. (2014) 2-Oxoglutarate: linking TCA cycle function with amino acid, glucosinolate, flavonoid, alkaloid, and gibberellin biosynthesis. *Front. Plant Sci.*, **5**, 552.

Bassel, G.W., Gaudinier, A., Brady, S.M., Hennig, L., Rhee, S.Y. and Smet, I. De (2012) Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. *Plant Cell*, **24**, 3859–3875.

Bénard, C., Bernillon, S., Biais, B., et al. (2015) Metabolomic profiling in tomato reveals diel compositional changes in fruit affected by source-sink relationships. *J. Exp. Bot.*, **66**, 3391–3404.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, **57**, 289–300.

Biais, B., Bénard, C., Beauvoit, B., et al. (2014) Remarkable reproducibility of enzyme activity profiles in tomato fruits grown under contrasting environments provides a roadmap for studies of fruit metabolism. *Plant Physiol.*, **164**, 1204–1221.

Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U. and Benfey, P.N. (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science*, **318**, 801–806.

Branlard, G., Dardevet, M., Saccomano, R., Lagoutte, F. and Gourdon, J. (2001) Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica*, **119**, 59–67.

Castle, S. and Randall, P. (1987) Effects of sulfur deficiency on the synthesis and accumulation of proteins in the developing wheat seed. *Aust. J. Plant Physiol.*, **14**, 503–516.

Chope, G., Wan, Y., Penson, S., Bhandari, D., Powers, S. and Shewry, P. (2014) Effects of genotype, season, and nitrogen nutrition on gene expression and protein accumulation in wheat grain. *J. Agric. Food Chem.*, **62**, 4399–4407.

Choulet, F., Alberti, A., Theil, S., et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science*, **345**, 1249721–1249721.

Dai, Z., Plessis, A., Vincent, J., et al. (2015) Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. *Plant J.*, **83**, 326–343.

Dhaka, V. and Khatkar, B.S. (2015) Effects of gliadin/glutenin and HMW-GS/LMW-GS ratio on dough rheological properties and bread-making potential of wheat varieties. *J. Food Qual.*, **38**, 71–82.

Don, C., Mann, G., Bekes, F. and Hamer, R.J. (2006) HMW-GS affect the properties of glutenin particles in GMP and thus flour quality. *J. Cereal Sci.*, **44**, 127–136.

Eriksen, J. (2009) Soil sulfur cycling in temperate agricultural systems. In *Advances in Agronomy*. pp. 55–89.

Franke, I., Resch, A., Dassler, T., Maier, T. and Bock, A. (2003) YfiK from Escherichia coli promotes export of O-acetylserine and cysteine. *J. Bacteriol.*, **185**, 1161–1166.

Fukushima, A., Kusano, M., Redestig, H., Arita, M. and Saito, K. (2009) Integrated omics approaches in plant systems biology. *Curr. Opin. Chem. Biol.*, **13**, 532–538.

Griffith, W. (1980) Determination of glutathione and glutathione disulfide glutathione reductase and 2-vinylpyridine. *Anal. Biochem.*, **212**, 207–212.

Hawkesford, M.J. (2000) Plant responses to sulphur deficiency and the genetic manipulation of sulphate transporters to improve S-utilization efficiency. *J. Exp. Bot.*, **51**, 131–138.

Hawkesford, M.J. and DeKok, L.J. (2006) Managing sulphur metabolism in plants. *Plant, Cell Environ.*, **29**, 382–395.

Hendriks, J.H.M., Kolbe, A., Gibon, Y., Stitt, M. and Geigenberger, P. (2003) ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species. *Plant Physiol.*, **133**.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T. and Saito, K. (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 10205–10210.

Hirel, B., LeGouis, J., Ney, B. and Gallais, A. (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative

genetics within integrated approaches. *J. Exp. Bot.*, **58**, 2369–2387.

Howarth, J.R., Parmar, S., Jones, J., et al. (2008) Co-ordinated expression of amino acid metabolism in response to N and S deficiency during wheat grain filling. *J. Exp. Bot.*, **59**, 3675–3689.

Husson, F., Josse, J., Le, S., Mazet, J. and & Husson, M.F. (2015) Package "FactoMineR".

Jamal, A., Moon, Y.S. and Abdin, M.Z. (2010) Sulphur -a general overview and interaction with nitrogen. *Aust. J. Crop Sci.*, **4**, 523–529.

Jelitto, T., Sonnewald, U., Willmitzer, L., Hajirezeai, M. and Stitt, M. (1992) Inorganic pyrophosphate content and metabolites in potato and tobacco plants expressing E. coli pyrophosphatase in their cytosol. *Planta*, **188**, 238–244.

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y. and Martinoia, E. (2011) Plant ABC transporters. *Arab. B.*, **9**, e0153.

Kichey, T., Hirel, B., Heumez, E., Dubois, F. and LeGouis, J. (2007) In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. *F. Crop. Res.*, **102**, 22–32.

Kidokoro, S., Watanabe, K., Ohori, T., et al. (2015) Soybean DREB1/CBF-type transcription factors function in heat and drought as well as cold stress-responsive gene expression. *Plant J.*, **81**, 505–518.

Kopriva, S. and Rennenberg, H. (2004) Control of sulphate assimilation and glutathione synthesis: interaction with N and C metabolism. *J. Exp. Bot.*, **55**, 1831–1842.

Kruse, C., Jost, R., Lipschis, M., Kopp, B., Hartmann, M. and Hell, R. (2007) Sulfur-enhanced defence: effects of sulfur metabolism, nitrogen supply, and pathogen lifestyle. *Plant Biol.*, **9**, 608–619.

Leroy, P., Guilhot, N., Sakai, H., et al. (2012) TriAnnot: a versatile and high performance pipeline for the automated annotation of plant genomes. *Front. Plant Sci.*, **3**, 1–14.

Likes, R., Madl, R.., Zeisel, S.. and Craig, S.A.. (2007) The betaine and choline content of a whole wheat flour compared to other mill streams. *J. Cereal Sci.*, **46**, 93–95.

Marcussen, T., Sandve, S.R., Heier, L., et al. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, **345**, 1250092–1250092.

Mewes, H.W. (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **30**, 31–34.

Moing, A., Maucourt, M., Renaud, C., et al. (2004) Quantitative metabolic profiling by 1-dimensional 1H-NMR analyses: Application to plant genetics and functional genomics. *Funct. Plant Biol.*, **31**, 889–902.

Moreno-Risueno, M.A., Busch, W. and Benfey, P.N. (2010) Omics meet networks—using systems approaches to infer regulatory networks in plants. *Curr. Opin. Plant Biol.*, **13**, 126–131.

Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H. and Hoefgen, R. (2003) Transcriptome analysis of sulfur depletion in Arabidopsis thaliana: interlacing of biosynthetic pathways provides response specificity. *Plant J.*, **33**, 633–50.

Nunes-Nesi, A., Carrari, F., Gibon, Y., et al. (2007) Deficiency of mitochondrial fumarase activity in tomato plants impairs photosynthesis via an effect on stomatal function. *Plant J.*, **50**, 1093–1106.

Palta, J. and Fillery, I. (1995) N application increases pre-anthesis contribution of dry matter to grain yield in wheat grown on a duplex soil. *Aust. J. Agric. Res.*, **46**, 519–531.

Pechanek, U., Karger, A., Gröger, S., Charvat, B., Schöggl, G. and Lelley, T. (1997) Effect of nitrogen fertilization on quantity of flour protein components, dough properties, and breadmaking quality of wheat. *Cereal Chem.*, **74**, 800–805.

Pont, C., Murat, F., Confolent, C., Balzergue, S. and Salse, J. (2011) RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum* L.). *Genome Biol.*, **12**, R18.

Proost, S., Bel, M. Van, Sterck, L., Billiau, K., Parys, T. Van, Peer, Y. Van de and Vandepoele, K. (2009) PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell*, **21**, 3718–3731.

Qin, D., Wang, F., Geng, X., Zhang, L., Yao, Y., Ni, Z., Peng, H. and Sun, Q. (2014) Overexpression of heat stress-responsive TaMBF1c, a wheat (*Triticum aestivum* L.) Multiprotein Bridging Factor, confers heat tolerance in both yeast and rice. *Plant Mol. Biol.*, **87**, 31–45.

R Core Team (2015) R: A language and environment for statistical computing. R foundation for

statistical computing, Vienna, Austria.

Randall, P., Spencer, K. and Freney, J. (1981) Sulfur and nitrogen fertilizer effects on wheat. I. Concentrations of sulfur and nitrogen and the nitrogen to sulfur ratio in grain, in relation to the yield response. *Aust. J. Agric. Res.*, **32**, 203.

Rausch, T. and Wachter, A. (2005) Sulfur metabolism: a versatile platform for launching defence operations. *Trends Plant Sci.*, **10**, 503–509.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.

Sanchez, R. and Zhou, M.-M. (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.*, **36**, 364–372.

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M. and Ligterink, W. (2016) Learning from co-expression networks: possibilities and challenges. *Front. Plant Sci.*, **7**, 444.

Shewry, P.R. and Halford, N.G. (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.*, **53**, 947–958.

Shewry, P.R., Tatham, A.S. and Halford, N.G. (2001) Nutritional control of storage protein synthesis in developing grain of wheat and barley. *Plant Growth Regul.*, **34**, 105–111.

Shimizu, M.M. and Mazzafera, P. (2000) A role for trigonelline during imbibition and germination of coffee seeds. *Plant Biol.*, **2**, 605–611.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Sosa-Valencia, G., Covarrubias, A.A. and Reyes, J.L. (2013) *Stress signaling in plants: genomics and proteomics perspective, Volume 1* M. Sarwat, A. Ahmad, and M. Abdin, eds., New York, NY: Springer New York.

Tea, I., Genter, T., Violleau, F. and Kleiber, D. (2005) Changes in the glutathione thiol-disulfide status in wheat grain by foliar sulphur fertilization : consequences for the rheological properties of dough. , **41**, 305–315.

Tisdale, S.L., Reneau, R.B. and Platou, J.S. (1986) Atlas of sulfur deficiencies. *Sulfur Agric.*, **27**, 295–314.

Tompkins, D. and Toffaletti, J. (1982) Enzymic determination of citrate in serum and urine , with use of the Worthington " Ultrafree " device. *Clin. Chem.*, **28**, 2031–2033.

Triboï, E., Martre, P. and Triboï-Blondel, A.-M. (2003) Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. *J. Exp. Bot.*, **54**, 1731–1742.

Urano, K., Kurihara, Y., Seki, M. and Shinozaki, K. (2010) "Omics" analyses of regulatory networks in plant abiotic stress responses. *Curr. Opin. Plant Biol.*, **13**, 132–138.

Västermark, Å., Almén, M., Simmen, M.W., Fredriksson, R. and Schiöth, H.B. (2011) Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster EamA (DUF6) before the radiation of Viridiplantae. *BMC Evol. Biol.*, **11**, 123.

Vincent, J., Martre, P., Gouriou, B., Ravel, C., Dai, Z., Petit, J.-M. and Pailloux, M. (2015) RulNet: A web-oriented platform for regulatory network inference, application to wheat –omics data. *PLoS One*, **10**, e0127127.

Wei, H., Yordanov, Y.S., Georgieva, T., Li, X. and Busov, V. (2013) Nitrogen deprivation promotes *Populus* root growth through global transcriptome reprogramming and activation of hierarchical genetic networks. *New Phytol.*, **200**, 483–497.

Wickham, H. (2011) ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.*, **3**, 180–185.

Wieser, H. (2007) Chemistry of gluten proteins. *Food Microbiol.*, **24**, 115–119.

Wieser, H., Gutser, R. and Tucher, S. von (2004) Influence of sulphur fertilisation on quantities and proportions of gluten protein types in wheat flour. *J. Cereal Sci.*, **40**, 239–244.

Wrigley, C., Cros, D., Archer, M., Downie, P. and Roxburgh, C. (1980) The sulfur content of wheat endosperm proteins and its relevance to grain quality. *Aust. J. Plant Physiol.*, **7**, 755–766.

Zhao, F., Hawkesford, M. and McGrath, S. (1999) Sulphur assimilation and effects on yield and quality of wheat. *J. Cereal Sci.*, **30**, 1–17.

Zörb, C., Grover, C., Steinfurth, D. and Hermann Mühling, K. (2010) Quantitative proteome analysis of wheat gluten as influenced by N and S nutrition. *Plant Soil*, **327**, 225–234.

Supplementary information

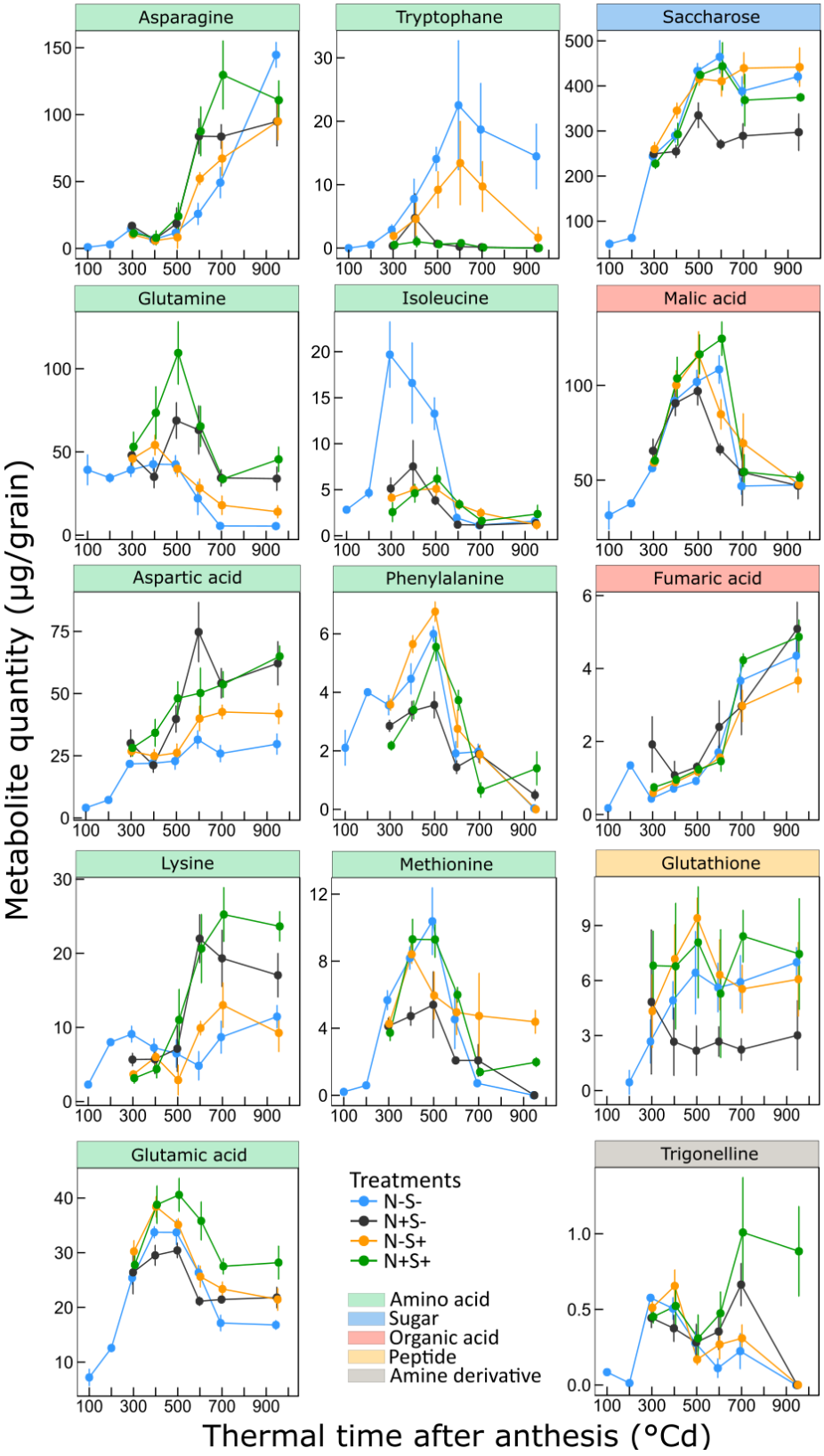


Figure S1. Mass per grain in the four N and S treatments of the 14 metabolites significantly impacted by nutrition during einkorn grain development. The four treatments were N-S- (0mM N and 0mM S), N+S- (6mM N and 0mM S), N-S+ (0.5mM N and 2mM S), N+S+ (6mM N and 2mM S), applied from 200 to 700°Cd after anthesis. Data are means ± 1 SE for n = 4 biological replicates.

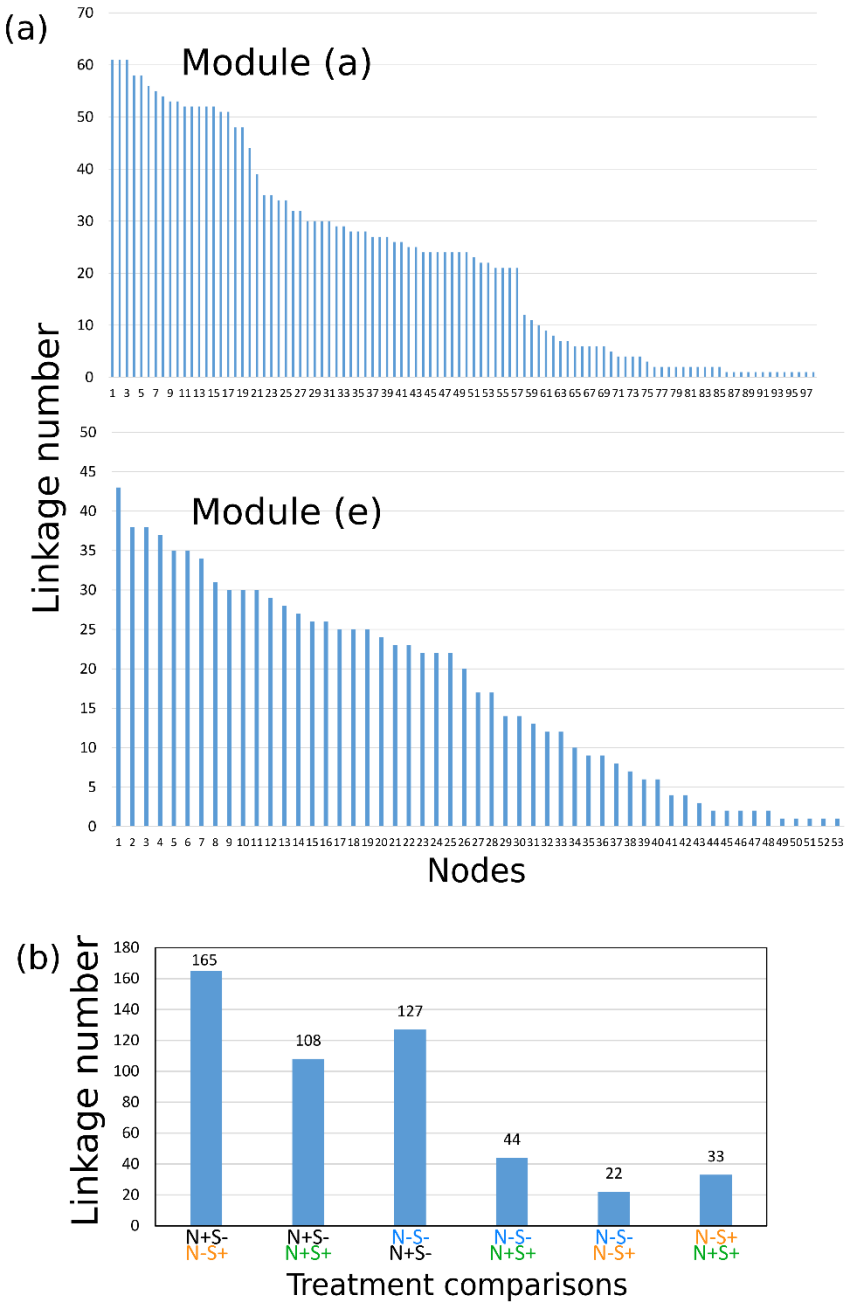


Figure S2. Linkage number in network analysis. (a) Number of edges associated to each attribute present in modules 1 and 5 in co-expression network (Figure 4). (b) Number of edges that were associated to each treatment comparison in the nutrition network (Figure 5).

Tables S1 and S2 are not necessary to understand the manuscript thesis since they correspond to quantitative data for metabolites and to gene expression data. Thus they have not been included in this manuscript.

Table S3. Summary of the 609 variables significantly impacted by N and S supply. Column 'Code' indicates codes used in figures 4, 5 and 6 to qualify transcripts (codes that begin with 'G' letter), metabolites and proteins (albumins-globulins and nuclear proteins are distinguished by codes that begin with 'Alg' and 'N', respectively). For transcriptomic data, Gene Ontology annotation of genes uncharacterized is given in parenthesis in column 'Functional annotation'. Sometimes, genes had no functional annotation. If the variable was part of the co-expression network, its module and number of associations are indicated in columns 'Module' and 'Degree', respectively. In columns related to the nutrition network, signs '+' or '-' indicate that the protein was present in network, '+' or '-' indicating a higher or a lower amount in the first treatment, respectively. In the last column, cluster number is indicated for transcriptomic data and refers to the clustering analysis presented in Figure 2.

Code	Functional annotation	Co-expression network		Nutrition network						Cluster (gene)
		Module	Degree	N+S- N-S-	N-S+ N-S-	N+S+ N-S-	N-S+ N+S-	N+S+ N+S-	N+S+ N-S+	
G1	Major intrinsic protein			+			-	-		2
G2	uncharacterized (GO: nucleic acid binding)	5	2							1
G3	Cysteine proteinase inhibitor									3
G4		5	22							1
G5	Xyloglucan endo-transglycosylase - XET - C-terminus					+			+	2
G6	1 -3-beta-glucan synthase component									1
G7	uncharacterized (GO: phospholipid binding)									1
G8	2OG-Fe- II - oxygenase superfamily						+		-	1
G9	Wound-induced protein									1
G10	Uncharacterised conserved protein - DUF2359									1
G11	uncharacterized (GO: peptidase inhibitor activity)							-		2
G12									+	3
G13	Cys-rich Gliadin N-terminal									1
G14	uncharacterized (GO: hydrolase activity)									3
G15	uncharacterized (GO: zinc ion binding)			+						1
G16	NAD binding domain of 6-phosphogluconate dehydrogenase									1
G17										1
G18				-			+	+		3
G19	Cys-rich Gliadin N-terminal	2	2		+		+	+		1
G20	uncharacterized (GO: transcription)	5	14							1
G21	Low-molecular-weight glutenin subunit	3	2				+	+		3
G22	Serine hydroxymethyltransferase	4	5	+			-	-		2
G23	uncharacterized (GO: metal ion binding)									1
G24										3
G25	uncharacterized (GO: cell periphery)									1
G26	uncharacterized (GO: transferase activity)									1
G27	Phosphoethanolamine N-methyltransferase 1					+				1
G28				+			-			2
G29	40S ribosomal protein S20									1
G30	Pollen proteins Ole e I like									1
G31	uncharacterized (GO: zinc ion binding)									1
G32	uncharacterized (GO: purine ribonucleoside binding)									1
G33	uncharacterized (GO: protein binding)									1
G34	uncharacterized (GO: primary metabolic process)	5	35							1
G35	Protein tyrosine kinase									3
G36	Bowman-Birk serine protease inhibitor family									3
G37	Maintenance of mitochondrial structure and function									3
G38	Gamma-gliadin	3	1	-				+		1
G39	uncharacterized (GO: mRNA processing)			+			-	-		1
G40	uncharacterized (GO: protein binding)			+			-	-		2
G41	uncharacterized (GO: transmembrane transport)									2
G42	Asparagine synthetase (Glutamine-hydrolyzing)			+						1
G43	Endosperm transfer cell specific PR60									1
G44	Basic endochitinase A	1	30	-						3
G45										2
G46	Allinase						-			2
G47	GHMP kinases C terminal									2
G48	Rhodanese-like domain									1
G49	Adenylyl-sulfate kinase			+			-	-		2
G50	Chitinase (Fragment)	1	34	-						3
G51	EamA-like transporter family	4	14	+			-	-		2
G52	unknown (GO: root morphogenesis)						-			3
G53	Glycosyl hydrolase family 1						-			1
G54	unknown			+			-	-		2
G55	unknown					+				1
G56	uncharacterized (GO: oxidation reduction process)	1	28	+						2

G57										1
G58	uncharacterized (GO: oxidation reduction process)			+			-	-		2
G59	Choline -ethanolamine kinase									1
G60	unknown (GO: zinc ion binding)			+			-	-		2
G61	Sulfate transporter N-terminal domain with GLY motif			+			-	-		2
G62	uncharacterized (GO: metal ion binding)									1
G63	uncharacterized (GO: protein binding)									3
G64										1
G65	DNA polymerase		1	+			-		+	1
G66							-			1
G67	uncharacterized (GO: defense response)									3
G68	Cytochrome P450	1	27							3
G69	uncharacterized (GO: integral component of membrane)		1				-			1
G70	uncharacterized (GO: cell wall macromolecule)					-				1
G71	uncharacterized (GO: protein binding)									1
G72	unknown (GO: defense response bacterium)	1	32	-						3
G73	uncharacterized (GO: protein binding)						-			1
G74	uncharacterized (GO: 4-alpha-glucanotransferase activity)									2
G75	uncharacterized (GO: binding)									1
G76	Beta purothionin			-			+	+		1
G77		1	23							3
G78	uncharacterized (GO: RNA glycosylase activity)			+						3
G79		5	26							1
G80										1
G81	Cupin	1	53							3
G82	Late embryogenesis abundant protein						+			1
G83	uncharacterized (GO: protein binding)									1
G84	Trehalose-phosphatase			+			-			2
G85	Fatty acid hydroxylase superfamily	1	22							3
G86	Catalase isozyme 1			-			+	+		1
G87	uncharacterized (GO: zinc ion binding)						-			3
G88	uncharacterized (GO: nucleic acid binding)									1
G89	Multiprotein bridging factor 1									1
G90	Glutathione S-transferase - N-terminal domain	4	1	+	-		-	-		2
G91	uncharacterized (GO: integral component of membrane)									2
G92										1
G93	HVA22-like protein	1	29	+						3
G94										1
G95	Possible lysine decarboxylase									1
G96	uncharacterized (GO: transferase activity)									1
G97	uncharacterized (GO: zinc ion binding)									3
G98	Eukaryotic cytochrome b561							-		2
G99	Serine carboxypeptidase			+			-			2
G100	NmrA-like family	4	3	+			-	-		2
G101										1
G102	Class II chitinase	1	34							3
G103	Bowman-Birk type proteinase inhibitor I-2B (Fragment)							+		3
G104	unknown									3
G105	uncharacterized (GO: oxidation reduction process)	1	21							3
G106	uncharacterized (GO: nutrient reservoir actitivy)									2
G107	unknown									3
G108	Tubulin C-terminal domain									1
G109	Alpha amylase - catalytic domain									1
G110	Defensin						+			1
G111	Stress responsive A -B Barrel Domain	1	24							3
G112				+			-		+	1
G113										1
G114	Protein BREAST CANCER SUSCEPTIBILITY 1-like protein				+					2
G115	Cysteine synthase			+			-	-		2
G116	Chitin recognition protein	1	52							3
G117	Chitinase class I	1	30							3
G118	uncharacterized (GO: killing of cells of other organism)									3
G119	Endoglucanase	5	43							1
G120	TPR repeat	4	2	+			-	-		2
G121	Chitin recognition protein	1	25	-						3
G122										1
G123	unknown	4	1	+			-	-		2
G124	uncharacterized (GO: killing of cells of other organism)					+		+		3
G125	unknown_function									3
G126	Non-specific lipid-transfer protein			-						3
G127	Rhodanese-like domain			+			-	-		2
G128										2
G129	uncharacterized (GO: heme binding)									1
G130	Ureide permease	1	24							3
G131	Hsp20 -alpha crystallin family					-				1
G132										3

G133							-		+	1
G134	unknown	5	23							1
G135	unknown									2
G136	uncharacterized (GO: zinc ion binding)						-			1
G137	Haloacid dehalogenase-like hydrolase									3
G138	E2F -DP family winged-helix DNA-binding domain									1
G139	Pectinacetylesterase	1	6							3
G140	unknown_function			+						2
G141	Glycosyl hydrolase family 1	1	30	-						3
G142				+			-	-		2
G143	60S acidic ribosomal protein P0									2
G144	Glycosyl hydrolase family 65 - N-terminal domain			-			+	+		1
G145	Sucrose synthase 2	5	34							1
G146	unknown (GO: regulation of amino acid export)			+		+				2
G147	Glycosyl hydrolase family 3 C-terminal domain									3
G148	Tubulin beta-1 chain	2	2	-	+		+	+		1
G149	Protein of unknown function - DUF506	4	3	+			-	-		2
G150	unknown									3
G151	unknown (GO: regulation of transcription from RNA polymerase II promoter)									1
G152	uncharacterized (GO: oxidation reduction process)					+				1
G153	uncharacterized (GO: viral process)									1
G154	uncharacterized (GO: primary metabolic process)	1	28							2
G155	Citrate synthase						+	+		3
G156	unknown									1
G157										3
G158	unknown									3
G159	unknown						-			1
G160	Late embryogenesis abundant protein	1	2	-						1
G161	unknown_function			+			-	-		2
G162	Aminotransferase class I and II									1
G163	uncharacterized (GO: protein binding)				-		-		+	1
G164	Probable allantoinase						+			1
G165	uncharacterized (GO: transcription)									3
G166						+		+		3
G167	Macrophage migration inhibitory factor - MIF						+	+		1
G168	uncharacterized (GO: cellular catabolic process)	1	22	-						3
G169	Glycosyltransferase like family 2	5	1							1
G170	uncharacterized (GO: purine ribonucleoside binding)								+	2
G171	uncharacterized (GO: integral component of membrane)									3
G172		5	38							1
G173	unknown (GO: lipid binding)	5	38							1
G174	RuBisCO large subunit (Fragment)									1
G175	unknown						-	-		2
G176				+			-	-		2
G177	uncharacterized (GO: zinc ion binding)						-			1
G178										1
G179										1
G180							-			1
G181	uncharacterized (GO: thiamine pyrophosphate binding)									2
G182		5	35	-						1
G183	MA3 domain			+			-		+	2
G184	Serpin - serine protease inhibitor			+			-	-		2
G185	uncharacterized (GO: oxidation reduction process)	4	2	+						2
G186	Cys-rich Gliadin N-terminal	3	3	-	+		+	+		1
G187										1
G188	uncharacterized (GO: phospholipid binding)									3
G189	Trypsin alpha-amylase inhibitor CMX1 CMX3				+		+			1
G190	unknown	1	26	-						3
G191				+						1
G192	unknown									3
G193					-		-			2
G194	uncharacterized (GO: purine ribonucleoside binding)						-			3
G195							-	-		1
G196	uncharacterized (GO: heat shock protein binding)									1
G197	unknown (GO: metal ion binding)									2
G198					+		+			1
G199						+	-		+	1
G200										1
G201	Aminotransferase class-III	4	4	+			-	-		1
G202	uncharacterized (GO: nucleic acid binding)	1	28						-	3
G203	unknown						-	-		1
G204	unknown	5	12							1
G205										1
G206	uncharacterized (GO: transcription)									1
G207	uncharacterized (GO: zinc ion binding)									1

G208	tRNA synthetase class II core domain - G - H - P - S and T					+				2
G209	Probable inositol oxygenase									2
G210	unknown			+				-		2
G211	DNA polymerase									3
G212	uncharacterized (GO: regulation of transcription from RNA polymerase II promoter)									1
G213				-				+		1
G214	uncharacterized (GO: transcription)									1
G215	Flavin containing amine oxidoreductase									3
G216										1
G217	uncharacterized (GO: binding)									1
G218	Phosphoadenosine phosphosulfate reductase family	4	3	+			-	-		2
G219										1
G220	OB-fold nucleic acid binding domain	5	26							1
G221					-				+	3
G222										1
G223	uncharacterized (GO: heat acclimatation)	1	1							3
G224	uncharacterized (GO: sucrose phosphate synthase activity)	1	27							3
G225				+			-			3
G226	ATP-sulfurylase	4	2	+			-	-		2
G227	Chitinase 2	1	1							3
G228	unknown									3
G229	Gamma-thionin family			+						1
G230	uncharacterized (GO: protein binding)									1
G231	unknown_function									3
G232										1
G233	Formate dehydrogenase mitochondrial	2	1	+			-	-		1
G234	Glycosyl hydrolases family 31			+			-	-		2
G235	DNA replication licensing factor MCM4	5	7							1
G236	Serine threonine-protein kinase AtPK19					+				1
G237	Peptidase family M28	1	1	-			+			3
G238	uncharacterized (GO: proteolysis)	5	13	+			-			1
G239	RWD domain						-			2
G240	ENTH domain							+	+	1
G241	Taurine catabolism dioxygenase TauD - TfdA family	4	2	+			-	-		2
G242	uncharacterized (GO: primary metabolic process)									1
G243								-		2
G244	uncharacterized (GO: zinc ion binding)									1
G245	Malonyl-CoA decarboxylase - MCD						-			1
G246	uncharacterized (GO: purine ribonucleoside binding)									1
G247	Sulfate transporter family			+			-	-		2
G248	uncharacterized (GO: cellular catabolic process)								-	3
G249	Cupin									1
G250	uncharacterized (GO: primary metabolic process)									1
G251	Diacylglycerol kinase	5	17							1
G252		5	25							1
G253	uncharacterized (GO: small GTPase mediated signal transduction)									1
G254	DNA replication licensing factor MCM6	5	22	+		+	-			1
G255	uncharacterized (GO: ATPase activity)									2
G256	unknown (GO : ATP binding)						-	-		3
G257	Xylanase inhibitor N-terminal			+			-	-		1
G258	EF-hand domain pair					-				1
G259	Sulfite exporter TauE -SafE									3
G260	Xylanase inhibitor N-terminal	5	20							1
G261										1
G262	Gamma-thionin family									1
G263	Myb-like DNA-binding domain					+				3
G264										3
G265							-			1
G266	uncharacterized (GO: polysaccharide biosynthetic process)									1
G267	uncharacterized (GO: ATP binding)						-			2
G268										1
G269	unknown	4	5	+			-	-		2
G270	Transmembrane amino acid transporter protein			+			-	-		2
G271	unknown (GO: lipid binding)	5	17							1
G272	Fructan 6-exohydrolase									1
G273										3
G274	uncharacterized (GO: oxidation reduction process)			+			-	-		2
G275	Putative 12-oxophytodienoate reductase 11									1
G276							-			1
G277	Betaine aldehyde dehydrogenase 2			+			-			2
G278	Putative xylanase inhibitor	5	2					+		1
G279	uncharacterized (GO: alpha-amylase activity)									3
G280				+						2
G281	Eukaryotic translation initiation factor 5A-4									1
G282	Rare lipoprotein A - RlpA --like double-psi beta-barrel					+				1

G283	uncharacterized (GO: binding)									3
G284	Kelch motif									2
G285	unknown									3
G286	Serpin - serine protease inhibitor	2	1	+			-	-		1
G287	Alcohol dehydrogenase	1	2	-			+	+		3
G288	Cyclin-dependent kinase inhibitor 4									3
G289	Auxin-repressed protein	1	51							3
G290	Potassium transporter									3
G291	PHD-finger			+			-	-		1
G292	Universal stress protein family			+						2
G293	Transmembrane amino acid transporter protein									1
G294	uncharacterized (GO: hydrolase activity)	1	12			+		+		3
G295	ABC transporter G family member 39					+			+	3
G296	Homocysteine S-methyltransferase	4	4	+			-	-		2
G297	Myb-like DNA-binding domain									3
G298	uncharacterized (GO: lipid binding)									1
G299	ACT domain			+			-	-		2
G300	uncharacterized (GO: protein binding)						-			1
G301	unknown (GO: ATP binding)									1
G302	Pyridine nucleotide-disulphide oxidoreductase									2
G303	Dihydroorotate dehydrogenase (quinone) mitochondrial									3
G304	unknown					+				2
G305	Glycerophosphoryl diester phosphodiesterase family									2
G306	uncharacterized (GO: protein binding)						-			1
G307	uncharacterized (GO: oxidation reduction process)	1	27							3
G308		5	24						+	1
G309	uncharacterized (killing of cells of other organism)					+				3
G310	unknown (GO: sulfate adenylyltransferase activity)	4	5	+		-		-		2
G311	Asparagine synthetase glutamine-hydrolyzing									3
G312										1
G313	uncharacterized (glucan biosynthetic process)	5	31							1
G314	uncharacterized (GO: zinc ion binding)								+	3
G315		5	25							1
G316	Puroindoline-A	3	2	-	+		+	+		3
G317										1
G318	HMGL-like									1
G319	uncharacterized (GO: zinc ion binding)									3
G320		1	24							3
G321	Raffinose synthase or seed imbibition protein Sip1					+				2
G322	uncharacterized (GO: nutrient reservoir activity)	2	3		+		+		-	1
G323	Avenin-like a2	2	3		+		+			1
G324	Senescence regulator									2
G325	uncharacterized (GO: calcium ion binding)									3
G326	Puroindoline-B									1
G327	Glutathione S-transferase - C-terminal domain	4	5	+			-	-		2
G328	Non-specific serine threonine protein kinase			+			-	-		2
G329	uncharacterized (GO: chromatin modification)									3
G330	Beta-1 3-glucanase	1	21							3
G331										3
G332	uncharacterized (GO: hydrolase activity)									3
G333										1
G334	WD domain - G-beta repeat			+						1
G335	uncharacterized (GO: signal peptide processing)									2
G336	uncharacterized (GO: structural molecule activity)									1
G337	uncharacterized (GO: purine NTP-dependent helicase activity)						-			3
G338	unknown (GO: ribonucleoside binding)	2	2	+			-	-		1
G339	unknown (GO: protein binding)									1
G340	Gamma-thionin family	5	3							1
G341	Plant PDR ABC transporter associated	1	24							2
G342	unknown_function			+			-	-		2
G343	uncharacterized (GO: protein binding)									2
G344										1
G345										1
G346	unknown (GO: serine-type endopeptidase inhibitor activity)	3	3	-	+		+	+		1
G347	uncharacterized (GO: protein binding)									1
G348										1
G349							+			1
G350	Nuclear transcription factor Y subunit C-2						-	-		2
G351										1
G352										3
G353	Tetrapyrrole - Corrin -Porphyrin - Methylases			+			-	-		2
G354	uncharacterized (GO: cation transport)									3
G355	Gamma-thionin family									1
G356	Glutathione S-transferase - C-terminal domain									3
G357	uncharacterized (GO: killing of cells of other organism)									3
G358	Purple acid phosphatase	1	35							3

G359	uncharacterized (GO: mRNA splicing)									3
G360	unknown_function									2
G361	unknown (GO: transcription factor activity)									2
G362										2
G363	uncharacterized (GO: nucleoside metabolic process)	5	4						-	1
G364	uncharacterized (GO: heme binding)									3
G365	RNA dependent RNA polymerase									3
G366	uncharacterized (GO: oxidation reduction process)			+						2
G367	uncharacterized (GO: oxidation reduction process)	1	2				+			3
G368	Beta-amylase	1	1	-			+	+		3
G369	Putative NADP-dependent oxidoreductase	4	4	+			-	-		2
G370	unknown					+		+		3
G371	uncharacterized (GO: nucleic acid binding)					-				2
G372										3
G373	unknown_function									1
G374							-			1
G375	unknown_function						-	-		2
G376		4	4				-	-		2
G377		5	23							1
G378	Peptide N-acetyl-beta-D-glucosaminyl asparaginase amidase A									2
G379	Tetratricopeptide repeat	4	4	+			-	-		2
G380	Defensin	3	1				+	+		1
G381	uncharacterized (GO: defense response)	1	39							3
G382	unknown									2
G383	Probable lipid transfer									1
G384	uncharacterized (GO: oxidation reduction process)									1
G385	uncharacterized (GO: multicellular organism development)	1	29							3
G386	BNR repeat-like domain						-			2
Asp	Aspartate		2			+				
Ile	Isoleucine			-	-	-				
Glu	Glutamate	5	12							
Mal	Malic acid					+				
Met	Methionine			-						
Phe	Phenylalanine			-			+		-	
Sacch	saccharose	5	28							
Trp	Tryptophane			-		-	+			
Gln	Glutamine					+				
Trig	Trigoneline									
Asn	Asparagine									
Lys	Lysine									
Fum	Fumarate									
GSH	Glutathion									
alpha	alpha-gliadin									
gamma	gamma-gliadin	1	2							
w1.2	omega-1,2-gliadin	1	25			+	-		+	
w5	omega-5-gliadin	1	24						+	
HMW	HMW-GS	1	21	+			-	-		
LMW	LMW-GS	1	21			+				
Alg1	Limit dextrinase type starch debranching enzyme	1	1	-			+	+		
Alg2	Alcohol dehydrogenase ADH1A	1	11	-			+	+		
Alg3	Puroindoline A			-		-	+			
Alg4	Triticin							+		
Alg5	GTPase SAR1									
Alg6	Subtilisin protease (Fragment)									
Alg7	Zeta-carotene desaturase			+			-	-		
Alg8	Gamma-gliadin						+			
Alg9	Gamma gliadin (Fragment)	1	1							
Alg10	Fructose-bisphosphate aldolase	5	9							
Alg11	Trypsin inhibitor			-			+			
Alg12	Aminotransferase	1	2							
Alg13	Purple acid phosphatase	1	6	+						
Alg14	Omega-gliadin (Fragment)	1	52				-			
Alg15	Protein disulfide isomerase family protein 4-1									
Alg16	RNA recognition domain containing protein,expressed									
Alg17	Protein phosphatase 2A structural subunit									
Alg18	High molecular weight glutenin 1Ax2.1	1	52							
Alg19	Starch branching enzyme IIa					-				
Alg20	Superoxide dismutase [Cu-Zn] (Fragment)									
Alg21	Translocon-associated protein subunit beta			+						
Alg22	Basic endochitinase C	1	52	-						
Alg23	Importin subunit alpha								+	
Alg24	Vicilin-like antimicrobial peptides 2-2	1	61							
Alg25	Xylulose kinase			-	-	-				
Alg26	Isocitrate dehydrogenase [NAD] catalytic subunit 5, mitochondrial									
Alg27	GrpE protein homolog									

Alg28	Putative aconitate hydratase, cytoplasmic						+			
Alg29	Glutaredoxin-C6	1	2							
Alg30	Aminopeptidase N	5	6							
Alg31	Acyl-[acyl-carrier-protein] desaturase 2, chloroplastic									
Alg32	zinc finger protein									
Alg33	Intracellular protease 1						+			
Alg34	Aspartic proteinase oryzasin-1	1	51							
Alg35	Chitinase 1	1	53							
Alg36	Ankyrin repeat domain-containing protein 2	1	58							
Alg37	DNA damage-inducible protein 1									
Alg38	transmembrane 214-B					-				
Alg39	amino acid selective channel protein									
Alg40	Putative NADP-dependent oxidoreductase P1	1	3			+				
Alg41	2-oxoglutarate dehydrogenase, mitochondrial	1	2							
Alg42	Peroxisomal multifunctional enzyme type 2									
Alg43	non-specific lipid transfer GPI-anchored 2-like									
Alg44	Glutaredoxin-C8	5	2				+			
Alg45	isoflavone reductase homolog irl-like	4	2	+			-	-		
Alg46	Cytochrome b5	1	6							
Alg47	Serpin-ZX	1	9					-		
Alg48	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha				-		-			
Alg49	Monothiol glutaredoxin-S11					+				
Alg50	Endogenous alpha-amylase/subtilisin inhibitor	1	2							
Alg51	Programmed cell death protein 4					+			+	
Alg52	Putative glutathione S-transferase GSTF1						+			
Alg53	2'-deoxymugineic-acid 2'-dioxygenase									
Alg54	Peptide methionine sulfoxide reductase A2-1									
Alg55	glutathionyl-hydroquinone reductase -like						-	-		
Alg56	Endoglucanase 11						+			
Alg57	proton pump-interactor 1-like									
Alg58	L-ascorbate peroxidase 1, cytosolic	5	9							
Alg59	Programmed cell death protein 4	1	4				-			
Alg60	Gamma-hordein-3				+		+	+		
Alg61	Bifunctional aspartokinase/homoserine dehydrogenase 2, chloroplastic									
Alg62	Pyruvate, phosphate dikinase 1, chloroplastic			-						
Alg63	Coatomer subunit gamma						-			
Alg64	Adenosine kinase 2	5	30							
Alg65	Sucrose synthase 2									
Alg66	Trypsin/alpha-amylase inhibitor CMX1/CMX3			-			+	+		
Alg67	UDP-glucuronic acid decarboxylase 1	5	6							
Alg68	ADP,ATP carrier protein, mitochondrial									
Alg69	Uncharacterized protein			+						
Alg70	Cell division protein ftsZ-like protein 2-1, chloroplastic								+	
Alg71	Clathrin heavy chain									
Alg72	Chitinase 2	1	61							
Alg73	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial									
Alg74	hypothetical protein TRIUR3_12434	5	30							
Alg75	Actin-7	1	7		-		-			
Alg76	DEAD-box ATP-dependent RNA helicase 37						-			
Alg77	Inositol-3-phosphate synthase									
Alg78	Avenin-3			-			+	+		
Alg79	Putative bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase E1			-			+	+		
Alg80	Protein argonaute 4A									
Alg81	caleosin 2	1	7							
Alg82	Beta-amylase	1	1	-			+	+		
Alg83	Starch branching enzyme 1	1	48							
Alg84	Glucose-1-phosphate adenylyltransferase large subunit, chloroplastic/amyloplastic									
Alg85	Phosphoglycerate kinase, chloroplastic	5	37							
Alg86	Cysteine synthase			+			-	-		
Alg87	Non-specific lipid-transfer protein 2P	1	6							
Alg88	Disproportionating enzyme									
Alg89	Starch branching enzyme IIb			-				+		
Alg90	Delta-1-pyrroline-5-carboxylate dehydrogenase						+			
Alg91	Ribulose bisphosphate carboxylase large chain	5	22							
Alg92	Expansin EXPB3						+			
Alg93	Glutamine synthetase									
Alg94	uncharacterized (GO: storage protein)			-			+	+		
Alg95	Glutathione transferase	1	30	-						
Alg96	Beta-D-glucan exohydrolase							-	-	
Alg97	Thaumatin-like protein	1	1	-				+		
Alg98	Low-molecular-weight glutenin subunit group 6 type IV	1	2	-			+	+		

Alg99	Starch branching enzyme 1	1	48							
Alg100	Glucose-1-phosphate adenylyltransferase									
Alg101	Tubulin beta-3 chain									
Alg102	glutathione s-transferase 3	5	1							
Alg103	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	1	1							
Alg104	probable aldo-keto reductase 1-like							+		
Alg105	Uncharacterized protein									
Alg106	importin subunit beta-1-like	5	8							
Alg107	12s seed storage globulin 1-like									
Alg108	stromal 70 kda heat shock-related chloroplastic									
Alg109	universal stress protein a-like protein									
Alg110	pyrophosphate-energized vacuolar membrane proton pump 1-like									
Alg111	Importin subunit alpha			+			-			
Alg112	clathrin light chain 2-like									
Alg113	glutamate--glyoxylate aminotransferase 2									
Alg114	ECERIFERUM 26-like									
Alg115	methylthioribose kinase						-	-		
Alg116	Xylose isomerase						+			
Alg117	chaperone protein chloroplastic-like						-		+	
Alg118	Uncharacterized protein								+	
Alg119	udp-glucuronic acid decarboxylase 2-like	5	4							
Alg120	Uncharacterized protein (Fragment)			+			-	-		
Alg121	carbonic anhydrase	1	2	-			+	+		
Alg122	flower-specific gamma-thionin precursor						+		-	
Alg123	uridine 5 -monophosphate synthase-like									
Alg124	cationic peroxidase SPC4-like	1	56	-						
Alg125	pectin acetylesterase 5-like	1	26							
Alg126	pyruvate decarboxylase	1	1				+	+		
Alg127	glutamate-rich wd repeat-containing protein 1									
Alg128	UDP-glucose 6-dehydrogenase			-						
Alg129	gibberellin 20 oxidase 2	4	5		-		-	-		
Alg130	rna recognition motif containing family protein	1	1							
Alg131	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	4	2							
Alg132	transport SEC23-like								+	
Alg133	vicilin-like antimicrobial peptides 2-2	1	58				+			
Alg134	cop9 signalosome complex subunit 4									
Alg135	Alcohol dehydrogenase-like 2	1	55							
Alg136	cop9 signalosome complex subunit 1									
Alg137	Uncharacterized protein									
Alg138	cbs domain-containing protein									
Alg139	probable n-succinyldiaminopimelate aminotransferase						-	-		
Alg140	probable sarcosine oxidase	1	54							
Alg141	probable sarcosine oxidase	1	44							
Alg142	ripening-related 1	1	35							
Alg143	prohibitin- mitochondrial-like									
Alg144	alpha,alpha-trehalose-phosphate synthase [UDP-forming] 6-like								+	
Alg145	Oleosin	1	61							
Alg146	rRNA N-glycosidase	1	6	+						
Alg147	acetolactate synthase small subunit 2, chloroplastic-like									
Alg148	hyoscyamine 6-dioxygenase									
Alg149	aspartic proteinase oryzasin-1-like									
Alg150	protein kinase superfamily protein				-		-			
Alg151	probable gamma-aminobutyrate transaminase 3, mitochondrial						-	-		
Alg152	Formate dehydrogenase, mitochondrial									
Alg153	fructokinase-2	5	2				+			
Alg154	Dolichyl-diphosphooligosaccharide-- glycosyltransferase 48 kDa subunit					-				
Alg155	sucrose synthase	5	29							
N1	Puroindoline A			-			+	+		
N2	RNA recognition domain containing protein,expressed	5	1							
N3	4S Ribosomal protein S15a									
N4	Heat shock protein 9									
N5	Importin subunit alpha									
N6	Alanine aminotransferase 2					+				
N7	Actin-depolymerizing factor 4									
N8	Histone H1						-			
N9	Crooked neck-like protein 1									
N10	Heat shock cognate 7 kDa protein 1									
N11	6S Ribosomal protein L18	1	1			-				
N12	DNA-directed RNA polymerases I, II, and III subunit RPABC1	5	25							
N13	dna-binding protein						-			
N14	14-3-3-like protein B									
N15	peter pan-like protein	5	10						+	
N16	DEAD-box ATP-dependent RNA helicase 27	5	14							
N17	phosphorylase	1	8							
N18	Histone H1						-			

N19	coiled-coil domain-containing protein 12	5	1							
N20	6S acidic Ribosomal protein P2B				+		+			
N21	6S Ribosomal protein L23a									
N22	Elongation factor 2									
N23	Multiprotein-bridging factor 1a	1	10							
N24	4S Ribosomal protein S11	1	4							
N25	Serine/arginine-rich splicing factor 4									
N26	phosphoglycerate kinase, cytolic	1	1							
N27	Poly(A)-binding protein									
N28	Eukaryotic translation initiation factor isoform 4G-1									
N29	Eukaryotic translation initiation factor 5A2									
N30	Histone H1 WH1B.1									
N31	tyrosine-protein phosphatase	5	30							
N32	fam10 family protein at4g22670-like									
N33	probable aldo-keto reductase 1	1	4							
N34	Importin subunit alpha									
N35	trihelix transcription factor			+						
N36	protein rrp5 homolog	5	2							
N37	DNA-directed RNA polymerase									
N38	guanine nucleotide-binding 3 homolog					+				
N39	protein rcc2									
N40	60s ribosomal protein l30					-				
N41	Spastin									
N42	formate--tetrahydrofolate ligase	1	5							
N43	dead-box atp-dependent rna helicase 28	5	1						+	
N44	brain acid soluble protein 1						+			
N45	nucleolar complex protein 3 homolog	5	27							
N46	betaine aldehyde dehydrogenase	1	4			+				
N47	Uncharacterized (GO: nucleic acid binding)									
N48	sucrose synthase							+		

3. CONCLUSIONS

L'analyse transcriptomique réalisée dans cette étude a révélé que 70% des gènes annotés sur le génome de *T. monococcum* étaient exprimés pendant le développement du grain. Récemment dans le grain de riz une proportion similaire a été observée (Biselli *et al.*, 2015). Parmi ces gènes exprimés, nous avons vu que 386 étaient très significativement différentiellement exprimés entre au moins deux traitements et à au moins un stade de développement donné pendant le remplissage du grain (entre 300 et 600°Cj). Beaucoup de ces gènes ont des fonctions liées à la mise en place des réserves et aux mécanismes de défense. Parmi les résultats, nous avons observé une diminution de l'expression de plusieurs gènes intervenant dans le métabolisme des sucres lorsqu'une forte quantité d'N est disponible pour le grain, appuyée par des résultats similaires observés sur les quantités de saccharose du grain. Le métabolisme des acides aminés se comporte de manière inverse et semble activé dans cette condition, comme le montrent l'expression plus forte de gènes liés à leur transport et à leur synthèse ainsi que la quantité totale d'acides aminés du grain. Parmi ces acides aminés, la glutamine et l'asparagine, qui sont les principaux acides aminés transportés des parties végétatives vers le grain (Masclaux-Daubresse *et al.*, 2010) ont été augmentés avec les traitements contenant des forts niveaux d'N. Les résultats obtenus dans cette étude confirment donc ceux obtenus dans le chapitre précédent (notamment le rôle du ratio N/S) et apportent des informations supplémentaires sur les gènes impliqués dans les mécanismes de réponse décrits.

L'analyse transcriptomique a clairement révélé un fort effet de la carence en S sur le niveau d'expression des gènes du grain de *T. monococcum*, carence provoquée par un apport d'N important sans ajout de S. Le clustering a permis d'identifier de nombreux gènes surexprimés dans cette condition, dont plusieurs sont apparus fortement co-exprimés dans les réseaux générés, supposant leur action synergique dans le grain. Parmi eux, outre des gènes impliqués dans le métabolisme des acides aminés mentionnés dans le paragraphe précédent, plusieurs gènes liés au transport et au métabolisme du S ont été retrouvés. L'activation du métabolisme du S en condition de carence est un mécanisme plusieurs fois décrit chez les plantes (Nakamura *et al.*, 1999 ; Lewandowska and Sirko, 2008). Certains gènes tels que des transporteurs pourraient avoir un rôle capital dans la réponse du grain à cette situation de carence. De plus, quelques facteurs de transcription étaient surexprimés dans cette condition et pourraient être impliqués dans le contrôle de ce mécanisme de régulation.

Si certains gènes ont pu être mis en évidence dans cette étude, l'interprétation des résultats demeure encore incomplète. En effet, plusieurs gènes n'ont pu être annotés fonctionnellement. Leur rôle putatif dans la réponse biologique peut être imaginé en observant les gènes caractérisés auxquels ils sont connectés au sein des réseaux de co-expression, sur la base du principe de « guilt by association » (Oliver, 2000). Néanmoins, il est prévu de poursuivre la caractérisation *in silico* de ces gènes afin d'obtenir le maximum d'informations sur les fonctions impactées dans le grain par l'N et le S.

Dans cette étude, nous avons proposé une interprétation de la réponse globale du grain de *T.monococcum* à la nutrition N et S. Toutefois, même si les interactions mises en évidence sont parfois complexes, il est nécessaire de garder à l'esprit que les conclusions tirées sont dépendantes des choix qui ont été faits (méthodes d'analyses, valeurs de seuils utilisées,...) et

une complexité beaucoup plus grande dans la réalité de la réponse cellulaire face au stress nutritionnel peut facilement être imaginée.

Références :

Biselli, C., Bagnaresi, P., Cavalluzzo, D., et al. (2015) Deep sequencing transcriptional fingerprinting of rice kernels for dissecting grain quality traits. *BMC Genomics*.

Lewandowska, M. and Sirko, A. (2008) Recent advances in understanding plant response to sulfur-deficiency stress. , **55**, 457–471.

Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L. and Suzuki, A. (2010) Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. *Ann. Bot.*, **105**, 1141–1157.

Nakamura, T., Yamaguchi, Y. and Sano, H. (1999) Four rice genes encoding cysteine synthase: isolation and differential responses to sulfur, nitrogen and light. *Gene*, **229**, 155–161.

Oliver, S. (2000) Guilt-by-association goes global. *Nature*, **403**, 601–603.

CHAPITRE 6 :
Conclusions générales et perspectives de la thèse

1. CONCLUSIONS

Cette thèse a apporté de la connaissance à la compréhension des mécanismes moléculaires pouvant réguler le développement du grain de blé et la mise en place de ses réserves en fonction de la nutrition. Le protéome nucléaire a occupé une place importante dans les analyses réalisées dans ce travail. Les protéines nucléaires identifiées ont tout d'abord permis de valider une méthode permettant leur extraction à partir de grains de blé. L'une des difficultés rencontrées pour ce type d'extraction a été la présence de PR au sein des extraits protéiques. Toutefois, d'un point de vue qualitatif, l'enrichissement en protéines nucléaires a été estimé entre 60 et 85% suivant le stade de développement des grains utilisés. Outre les protéines histones et protéines ribosomales, attendues et retrouvées en abondance au sein des extraits nucléaires, un nombre important de protéines intervenant dans divers processus ont pu être identifiées, jusqu'à 1677 retrouvées dans le grain de *T. monococcum*. Cette liste de protéines apporte des informations sur le fonctionnement du noyau, qui reste aujourd'hui peu connu chez les plantes malgré son rôle central dans l'organisation et l'expression du génome végétal (Petrovská *et al.*, 2015).

L'analyse de ce sous-protéome chez le blé tendre et l'engrain a permis de mettre en évidence des protéines qui pourraient avoir un rôle important dans la régulation de la transcription. Chez le blé tendre, des protéines liées à l'organisation de la chromatine (e.g. complexe de remodelage de la chromatine, protéines HMG, Histone déacétylase) ont présenté des profils d'accumulation suggérant des modifications dans la conformation de l'ADN en fin de phase de cellularisation et au début de la phase de remplissage. Chez l'engrain nous avons pu voir que ces mécanismes épigénétiques, *via* la méthylation de l'ADN, pourraient avoir un rôle dans la réponse du grain à un changement de nutrition au cours du remplissage, notamment suite à un apport de S. Comme il l'a été discuté dans le chapitre 4, la méthylation des promoteurs de gènes de PR observée chez l'orge semble être impliquée dans la régulation de leur transcription et dans le contrôle du développement du grain (Sørensen *et al.*, 1996; Radchuk *et al.*, 2005). La méthylation de l'ADN pourrait également modifier la capacité de liaison des FT à leur séquence promotrice cible, comme il l'a été observé pour le FT *Opaque2* chez le maïs (Sturaro and Viotti, 2001 ; Locatelli *et al.*, 2009).

Dans l'ensemble des jeux de données du protéome nucléaire, peu de FT, du moins fonctionnellement annotés de la sorte, ont été identifiés et quantifiés. De plus, parmi les protéines répondant à la nutrition, il était attendu de retrouver les FT du modèle de régulation transcriptionnelle établi chez l'orge et qui sont également présents chez les céréales (Rubio-Somoza *et al.*, 2006 ; Verdier and Thompson, 2008). Toutefois l'un d'eux, PBF, a été identifié dans le grain de *T. monococcum* au stade 300°Cj après floraison. D'après les données de transcriptomique obtenues à partir du même matériel biologique, il est apparu que ce FT était le plus exprimé des huit FT du modèle de régulation et présentait un pic d'expression au stade 300°Cj (Résultats non présentés). Ceci suggère que l'absence de ces FT des jeux de données du protéome nucléaire, pourtant conséquents, résulte d'une trop faible abondance de ces protéines régulatrices. Malgré un fractionnement subcellulaire, probablement que cette faible quantité limite techniquement leur détection au sein de mélanges peptidiques complexes, composés de nombreux peptides issus de protéines majoritaires telles que les histones, limitant ainsi leur identification par spectrométrie de masse.

Néanmoins, plusieurs protéines possédant un domaine de liaison à l'ADN ont été significativement impactées par la nutrition (21% des protéines nucléaires répondant à la nutrition). L'apport de N et de S impacte donc des régulateurs transcriptionnels, qui pourraient jouer un rôle dans la régulation du développement du grain et la mise en place de ses réserves.

Les différents traitements appliqués à la culture de *T. monococcum* ont conduit à différents rapports N/S du grain (Figure 25a) dès 300°Cj après floraison, au début de la période de remplissage. Ces différents ratios N/S se traduisent à maturité par différentes compositions en PR du grain, les PR riches en S étant plus fortement accumulées avec un apport de S et les PR pauvres en S étant fortement accumulées en présence d'N, confirmant ainsi les résultats précédemment obtenus chez le blé tendre (Wieser *et al.*, 2004 ; Zörb *et al.*, 2010). Ce rapport N/S montre une carence en S dans deux des quatre traitements appliqués (N-S- et N+S-). Il est très fortement augmenté avec le traitement N+S-, c'est-à-dire lorsque de l'N a été apporté à la culture sans apport concomitant de S, conduisant à une situation de forte carence en S dans le grain. Dans cette condition, la composition du grain en PR est clairement impactée avec une augmentation nette des HMW-GS au détriment de certaines gliadines (Figure 25b). Ces résultats montrent une nouvelle fois l'importance d'un apport conjoint d'N et de S. En revanche, lorsque le S est apporté en grande quantité sans apport de N, le ratio N/S, bien qu'il soit le plus faible dans cette condition de nutrition, n'est pas aussi nettement impacté. Il semble donc que dans cette condition, l'N du grain soit issu en grande partie de la remobilisation de l'N accumulé dans les parties aériennes pendant le développement végétatif de la plante (Kichey *et al.*, 2007). La situation de carence en S observée lorsque l'N est apporté seul traduit probablement une faible remobilisation du S. En effet, chez le blé, le S du grain est issu pour moins de 50% de la remobilisation (Monaghan *et al.*, 1999).

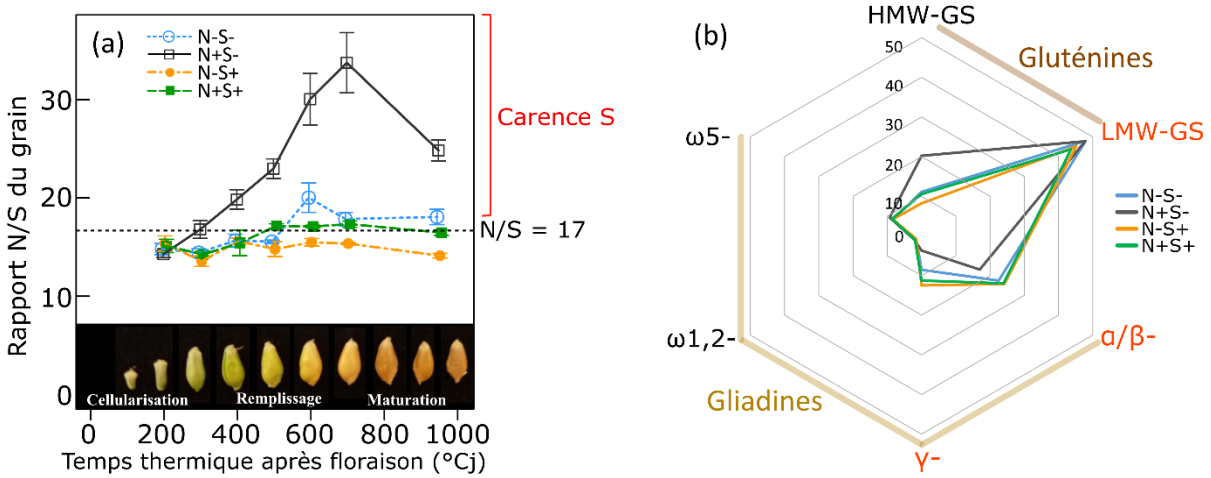


Figure 25. Effets de la nutrition azotée et soufrée sur le rapport N/S du grain (a) et sur la composition du grain mature chez T. monococcum (b). En (b) les données sont exprimées en pourcentage de la quantité totale de PR mesurée en $g.g^{-1}$ de matière sèche.

L'apport de S a conduit à une quantité plus importante de PR, due à une forte accumulation des gliadines, essentiellement les sous-unités α/β- et γ- dont les vitesses et durée d'accumulation ont été nettement augmentées. Nous avons ainsi pu observer que chez l'engrain, un apport de S après floraison conduit à une quantité plus importante de PR, sans vraiment avoir

une incidence sur la composition du grain, probablement du fait d'une bonne remobilisation de l'N. La plus grande quantité de PR a été observée lorsque de l'N et du S ont été apportés de manière conjointe et ce résultat est sûrement lié à la plus forte abondance observée pour plusieurs protéines nucléaires associées à la transcription et la maturation des ARNm.

Contrairement à ce que l'on aurait pu attendre, la majorité des ARNs, protéines, et métabolites quantifiés n'a pas répondu de cette façon-là à la nutrition. L'hypothèse principale est que ces variables ne répondent pas à la quantité totale d'N et de S dans le grain mais plutôt à la modification du rapport N/S. En effet, les perturbations les plus importantes ont été obtenues en réponse à un apport de N ou de S seul. Ainsi, nous avons observé une forte expression et accumulation des gènes et protéines impliqués dans le métabolisme et le transport du S, sans doute pour compenser la situation de carence provoquée par un fort apport de N sans ajout de S. Dans le même temps, de nombreux ARNs et protéines liés au métabolisme des acides aminés ont vu leur quantité augmenter face à un fort apport d'N. Ces observations ont été confortées par des résultats similaires obtenus au niveau des acides aminés quantifiés. De nombreuses variables liées au transport ont également été activées par l'N. Il s'agit de transporteurs de sulfate et d'acides aminés ou encore d'importines impliquées dans le transport nucléo-cytoplasmique. Certains de ces transporteurs pourraient avoir un rôle central dans l'ajustement du métabolisme du grain en réponse à une modification du rapport N/S, comme par exemple le transporteur EamA-like qui représentait un « hub » au sein du module de réponse à l'N dans le réseau de co-expression.

En parallèle de ces changements observés au niveau du métabolisme et du transport cellulaire, comme nous l'avons vu précédemment plusieurs gènes ou protéines possédant des domaines de liaison à l'ADN ont montré des variations significatives en réponse à la nutrition. Pour ces variables, il est difficile de définir un type majoritaire de réponse à la nutrition, toutes étant retrouvées à des endroits très différents au sein des réseaux biologiques générés. Plusieurs hypothèses peuvent être émises pour l'expliquer. Peut-être que ces protéines agissent seules, chacune ayant un rôle associé à une situation précise de stress. Peut-être qu'au contraire ces protéines forment des complexes, mais dont les partenaires n'ont pu être mis en évidence dans cette étude du fait de leur quantité très faible dans la cellule. Quoi qu'il en soit ces gènes pourraient agir comme des senseurs du rapport N/S ou de la quantité totale d'N et de S et avoir un rôle très important dans l'adaptation du grain pour l'accumulation de ses réserves face à un stress nutritionnel.

Si les FT connus pour être impliqués dans la régulation de la synthèse des PR n'ont pu être détectés en protéomique, leur niveau d'expression qui a pu être quantifié par RNA-Seq n'a pas été affecté par la nutrition. Ces gènes ne semblent donc pas régulés par l'N et le S au niveau transcriptionnel, du moins dans les conditions d'analyse que nous avons utilisées. On peut alors émettre l'hypothèse que ces FT pourraient être régulés par la nutrition à d'autres niveaux, par exemple par des modifications post-traductionnelles. On sait en effet que ces modifications peuvent affecter la conformation, l'activité ou encore la localisation des protéines. Chez le maïs par exemple, il a été démontré que l'état de phosphorylation d'Opaque 2 influence sa capacité de liaison aux promoteurs de gènes de PR (Ciceri *et al.*, 1997). La phosphorylation est une des modifications post-traductionnelles les plus caractérisées et est impliquée dans de nombreuses réponses physiologiques (Bond *et al.*, 2011 ; Ytterberg and Jensen, 2010). Chez *A. thaliana*, il

a été démontré que plusieurs régulateurs transcriptionnels subissent une modification de leur état de phosphorylation suite à un apport d'N (Engelsberger and Schulze, 2012).

Nous avons pu visualiser chez le blé tendre via des colorations de gels 2D au Pro-Q Diamond (un colorant spécifique des groupements phosphates des phosphopeptides) que certaines protéines nucléaires étaient phosphorylées en fin de phase de cellularisation. Il s'agit entre autres d'une protéine HMG et une histone déacétylase qui ont un rôle dans l'organisation de la chromatine. Ces analyses n'ont pas été poursuivies. Néanmoins, elles appuient l'hypothèse que la phosphorylation pourrait jouer un rôle majeur dans la régulation du développement du grain. De plus, l'impact de l'N et du S sur la quantité de plusieurs kinases et phosphatases sous-entend l'implication de mécanismes de phosphorylation/déphosphorylation dans la réponse du grain à la nutrition.

2. PERSPECTIVES

Les travaux réalisés au cours de cette thèse ont apporté de nouvelles informations sur la nature et les fonctions des protéines nucléaires de la cellule végétale et sur les mécanismes moléculaires mis en jeu en réponse à la nutrition N et S au cours du développement du grain de blé.

En intégrant différents types de données -omiques, nous avons pu voir que la réponse du grain à la nutrition met en jeu de nombreux gènes, protéines et métabolites, présentant des profils très variés, témoignant ainsi de la complexité de la réponse cellulaire. Bien sûr, la vision globale obtenue ne rend pas compte de l'intégralité des niveaux de régulation impliqués dans ce type de réponse. La cellule réagit probablement de manière bien plus complexe que ce qui a pu être observé. En effet, comme nous l'avons vu les modifications post-traductionnelles et les mécanismes épigénétiques pourraient tenir une place importante dans les processus de régulation mis en jeu. Des approches d'étude du phosphoprotéome et de la méthylation de l'ADN pourraient être envisagées pour compléter notre compréhension des effets de l'N et du S sur le grain de blé. Ces éléments permettraient d'apporter des pièces importantes au puzzle de la régulation de l'accumulation des PR du grain (Mehrotra *et al.*, 2009).

Au sein des réseaux générés dans cette thèse, les liens entre attributs reflètent leur co-expression ou co-accumulation. Les réseaux sont donc non-orientés. Pour aller plus loin dans la compréhension des mécanismes de régulation de l'accumulation des réserves du grain, il serait intéressant d'identifier le sens de ces interactions. Tout d'abord, les interactions protéine-ADN et protéine-protéine qui ont été démontrées expérimentalement pourraient être ajoutées aux réseaux obtenus. Puis dans cette optique, certains gènes apparus comme des « hub » au sein des réseaux pourrait faire l'objet d'études plus fonctionnelles. Une façon de procéder consisterait à inspecter la variabilité génétique existante pour ces gènes centraux. Si du polymorphisme de séquence est détecté, la culture des variétés de blé concernées et l'analyse des variations d'expression de ces gènes et des gènes avec qui ils sont fortement co-exprimés dans les réseaux permettrait de détecter des locus de traits quantitatifs d'expression (eQTL ; Kliebenstein *et al.*, 2006 ; Kliebenstein, 2009). D'une part, en visualisant l'influence des

variations d'expression d'un gène central sur l'expression d'autres gènes, le sens de certaines interactions existant dans les réseaux serait déterminé et les liens de régulation génique seraient alors mis en évidence. D'autre part, une analyse des PR en parallèle permettrait alors de relier les variations quantitatives des gènes étudiés au phénotype d'intérêt, à savoir la composition du grain du blé.

Nos travaux se sont inscrits dans un sujet de recherche à connotation très fondamentale. En effet, il faut garder à l'esprit que les résultats obtenus sont issus de culture de blé en conditions contrôlées. Les dispositifs choisis nous ont permis de contrôler au mieux la disponibilité de l'N et du S au sein du substrat et éviter l'effet d'autres variables environnementales. Nos résultats auraient pu être différents si les plantes avaient été cultivées en plein champ, où la disponibilité en éléments nutritifs est largement dépendante de la compaction du sol ou encore de la pluviométrie. Toutefois, les résultats obtenus suggèrent des pistes de travail utilisables en sélection variétale à moyen terme. Certains gènes mis en évidence, notamment des transporteurs activés en situation de carence en S, ou encore des régulateurs transcriptionnels, pourraient faire l'objet d'études fonctionnelles et pourquoi pas, s'avérer utiles dans l'objectif de maitrise de la qualité du grain de blé dans un contexte de limitation des intrants.

Références :

Bond, A.E., Row, P.E. and Dudley, E. (2011) Post-translation modification of proteins; methodologies and applications in plant sciences. *Phytochemistry*, **72**, 975–996.

Ciceri, P., Gianazza, E., Lazzari, B., Lippoli, G., Genga, A., Hoschek, G., Schmidt, R.J. and Viotti, A. (1997) Phosphorylation of Opaque2 changes diurnally and impacts its DNA binding activity. , **9**, 97–108.

Engelsberger, W.R. and Schulze, W.X. (2012) Nitrate and ammonium lead to distinct global dynamic phosphorylation patterns when resupplied to nitrogen-starved Arabidopsis seedlings. *Plant J.*, **69**, 978–95.

Kichey, T., Hirel, B., Heumez, E., Dubois, F. and LeGouis, J. (2007) In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. *F. Crop. Res.*, **102**, 22–32.

Kliebenstein, D. (2009) Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annu. Rev. Plant Biol.*, **60**, 93–114.

Kliebenstein, D.J., West, M. AL, Leeuwen, H. van, Loudet, O., Doerge, R. and St Clair, D. a (2006) Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics*, **7**, 308.

Locatelli, S., Piatti, P., Motto, M. and Rossi, V. (2009) Chromatin and DNA modifications in the Opaque2-mediated regulation of gene transcription during maize endosperm development. *Plant Cell*, **21**, 1410–1427.

Mehrotra, R., Kumar, S., Mehrotra, S. and Singh, B.D. (2009) Seed storage protein gene regulation- A jig-saw puzzle. *Indian J. Biotechnol.*, **8**, 147–158.

Monaghan, J.M., Scrimgeour, C.M., Stein, W.M., Zhao, F.J. and Evans, E.J. (1999) Sulphur accumulation and redistribution in wheat (*Triticum aestivum*): a study using stable sulphur isotope ratios as a tracer system. *Plant, Cell Environ.*, **22**, 831–839.

Petrovská, B., Šebela, M. and Doležel, J. (2015) Inside a plant nucleus: discovering the proteins. *J. Exp. Bot.*, **66**, 1627–1640.

Radchuk, V. V., Sreenivasulu, N., Radchuk, R.I., Wobus, U. and Weschke, W. (2005) The methylation cycle and its possible functions in barley endosperm development. *Plant Mol. Biol.*, **59**, 289–307.

Rubio-Somoza, I., Martinez, M., Abraham, Z., Diaz, I. and Carbonero, P. (2006) Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. *Plant J.*, **47**, 269–281.

Sørensen, M.B., Müller, M., Skerritt, J. and Simpson, D. (1996) Hordein promoter methylation and transcriptional activity in wild-type and mutant barley endosperm. *Mol. Gen. Genet.*, **250**, 750–760.

Sturaro, M. and Viotti, A. (2001) Methylation of the Opaque2 box in zein genes is parent-dependent and affects O2 DNA binding activity in vitro. *Plant Mol. Biol.*, **46**, 549–560.

Verdier, J. and Thompson, R.D. (2008) Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.*, **49**, 1263–1271.

Wieser, H., Gutser, R. and Tucher, S. von (2004) Influence of sulphur fertilisation on quantities and proportions of gluten protein types in wheat flour. *J. Cereal Sci.*, **40**, 239–244.

Ytterberg, a J. and Jensen, O.N. (2010) Modification-specific proteomics in plant biology. *J. Proteomics*, **73**, 2249–66.

Zörb, C., Grover, C., Steinfurth, D. and Hermann Mühling, K. (2010) Quantitative proteome analysis of wheat gluten as influenced by N and S nutrition. *Plant Soil*, **327**, 225–234.

RÉSUMÉ

L'augmentation des rendements est un enjeu majeur chez les céréales. Dans cet objectif, il est nécessaire de maintenir la qualité du grain de blé, qui est principalement déterminée par sa teneur et sa composition en protéines de réserve. En effet, une forte relation négative existe entre le rendement et la teneur en protéines. Par ailleurs, la qualité du grain est fortement influencée par la disponibilité en azote et en soufre dans le sol. La limitation des apports d'intrants azotés à la culture et la carence en soufre récemment observée dans les sols représentent ainsi des difficultés supplémentaires pour maitriser cette qualité. Une meilleure connaissance des mécanismes moléculaires impliqués dans le contrôle du développement du grain et la mise en place de ses réserves protéiques en réponse à la nutrition azotée et soufrée est donc primordiale. L'objectif de cette thèse a ainsi été d'apporter de nouveaux éléments à la compréhension de ces processus de régulation, aujourd'hui peu connus. Pour cela, les approches -omiques sont apparues comme une stratégie de choix pour identifier les acteurs moléculaires mis en jeu. Le protéome nucléaire a été une cible importante dans les travaux menés. L'étude de ces protéines nucléaires a révélé certains régulateurs transcriptionnels qui pourraient être impliqués dans le contrôle de la mise en place des réserves du grain. Dans une approche combinant des données de protéomique, transcriptomique et métabolomique, une vision intégrative de la réponse du grain à la nutrition azotée et soufrée a été obtenue. L'importance d'un apport de soufre dans le contrôle de la balance azote/soufre du grain, déterminante pour la composition du grain en protéines de réserve, a été clairement vérifiée. Parmi les changements observés au niveau du métabolisme cellulaire, certains des gènes affectés par la modification de cette balance pourraient orchestrer l'ajustement de la composition du grain face à des situations de carences nutritionnelles. Ces nouvelles connaissances devraient permettre de mieux maitriser la qualité du grain de blé dans un contexte d'agriculture durable.

Mots clés : Blé, grain, protéines de réserve, azote, soufre, protéines nucléaires, données omiques, réseaux biologiques

ABSTRACT

Improving the yield potential of cereals represents a major challenge. In this context, wheat grain quality has to be maintained. Indeed, grain quality is mainly determined by the content and the composition of storage proteins, but there is a strongly negative correlation between yield and grain protein concentration. In addition, grain quality is strongly influenced by the availability of nitrogen and sulfur in soils. Nowadays, the limitation of nitrogen inputs, and also the sulfur deficiency recently observed in soils represent major difficulties to control the quality. Therefore, understanding of molecular mechanisms controlling grain development and accumulation of storage proteins in response to nitrogen and sulfur supply is a major issue. The objective of this thesis was to create knowledge on the comprehension of these regulatory mechanisms. For this purpose, the best strategy to identify molecular actors involved in these processes consisted of -omics approaches. In our studies, the nuclear proteome was an important target. Among these proteins, we revealed some transcriptional regulators likely to be involved in the control of the accumulation of grain storage compounds. Using an approach combining proteomic, transcriptomic and metabolomic data, the characterization of the integrative grain response to the nitrogen and sulfur supply was obtained. Besides, our studies clearly confirmed the major influence of sulfur in the control of the nitrogen/sulfur balance that determines the grain storage protein composition. Among the changes observed in the cell metabolism, some genes were disturbed by the modification of this balance. Thus these genes could coordinate the adjustment of grain composition in response to nutritional deficiencies. These new results contribute in facing the challenge of maintaining wheat grain quality with sustainable agriculture.

Keywords: Wheat, grain, storage proteins, nitrogen, sulfur, nuclear protein, -omics data, biological network