



Proteomic Data Integration Highlights Central Actors Involved in Einkorn (*Triticum monococcum* ssp. *monococcum*) Grain Filling in Relation to Grain Storage Protein Composition

OPEN ACCESS

Edited by:

Antonio Masi,
University of Padua, Italy

Reviewed by:

Venkatesh Periyakavanam
Thirumalaikumar,
Max Planck Institute of Molecular
Plant Physiology, Germany
Carlos Alberto Labate,
University of São Paulo, Brazil

***Correspondence:**

Emmanuelle Bancel
emmanuelle.bancel@inra.fr

† Present address:

Titouan Bonnot,
Department of Botany and Plant
Sciences, University of California,
Riverside, Riverside, CA,
United States
Pierre Martre,
LEPSE, INRA, Montpellier SupAgro,
Université de Montpellier, Montpellier,
France

Specialty section:

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

Received: 15 February 2019

Accepted: 07 June 2019

Published: 04 July 2019

Citation:

Bancel E, Bonnot T, Davanture M,
Alvarez D, Zivy M, Martre P, Déjean S
and Ravel C (2019) Proteomic Data
Integration Highlights Central Actors
Involved in Einkorn (*Triticum*
monococcum ssp. *monococcum*)
Grain Filling in Relation to Grain
Storage Protein Composition.
Front. Plant Sci. 10:832.
doi: 10.3389/fpls.2019.00832

Emmanuelle Bancel^{1,2*}, Titouan Bonnot^{1,2†}, Marlène Davanture³, David Alvarez^{1,2}, Michel Zivy³, Pierre Martre^{1,2†}, Sébastien Déjean⁴ and Catherine Ravel^{1,2}

¹ UMR GDEC, Institut National de la Recherche Agronomique (INRA), Université Clermont Auvergne, Clermont-Ferrand, France, ² UMR1095, Genetics Diversity and Ecophysiology of Cereals, Clermont Auvergne University, Clermont-Ferrand, France, ³ UMR GQE, Institut National de la Recherche Agronomique (INRA), Centre National de la Recherche Scientifique (CNRS), Agro ParisTech, Université Paris-Sud – Université Paris-Saclay, Gif-sur-Yvette, France, ⁴ Institut de Mathématiques de Toulouse, UMR5219 Université de Toulouse, Centre National de la Recherche Scientifique (CNRS), Toulouse, France

Albumins and globulins (AGs) of wheat endosperm represent about 20% of total grain proteins. Some of these physiologically active proteins can influence the synthesis of storage proteins (SPs) (gliadins and glutenins) and consequently, rheological properties of wheat flour and processing. To identify such AGs, data, (published by Bonnot et al., 2017) concerning abundance in 352 AGs and in the different seed SPs during grain filling and in response to different nitrogen (N) and sulfur (S) supply, were integrated with mixOmics R package. Relationships between AGs and SPs were first unraveled using the unsupervised method sparse Partial Least Square, also known as Projection to Latent Structure (sPLS). Then, data were integrated using a supervised approach taking into account the nutrition and the grain developmental stage. We used the block.splda procedure also referred to as DIABLO (Data Integration Analysis for Biomarker discovery using Latent variable approaches for Omics studies). These approaches led to the identification of discriminant and highly correlated features from the two datasets (AGs and SPs) which are not necessarily differentially expressed during seed development or in response to N or S supply. Eighteen AGs were correlated with the quantity of SPs per grain. A statistical validation of these proteins by genetic association analysis confirmed that 5 out of this AG set were robust candidate proteins able to modulate the seed SP synthesis. In conclusion, this latter result confirmed that the integrative strategy is an adequate way to reduce the number of potentially relevant AGs for further functional validation.

Keywords: albumin-globulin, data integration, nitrogen, proteomic, grain development, storage protein, sulfur, *Triticum monococcum*

INTRODUCTION

The total proteins in cereal grains mainly consist of grain storage proteins (SPs) and metabolic proteins. In wheat, SPs gliadins and glutenins, make up 60–80% of the total protein content. Glutenins are classified as high-molecular-weight or low-molecular-weight glutenin subunits (HMW-GS and LMW-GS, respectively) and gliadins are divided into α -, γ -, and ω -gliadins. Grain SPs have been extensively studied in contrast to metabolic proteins, which have not yet been well characterized. Metabolic proteins correspond to the soluble albumin-globulin (AG) fraction, which contributes to 15–20% of the total grain protein content in bread wheat (Shewry and Halford, 2002). The effect of developmental changes in the metabolic proteins of wheat endosperm was first investigated using two-dimensional gel electrophoresis, image analysis and mass spectrometry (2-DE/MS/proteomics; Vensel et al., 2005; Merlino et al., 2009; Debiton et al., 2011; Tasleem-Tahir et al., 2012). Recently, several quality-related albumins and globulins were identified in an elite Chinese wheat cultivar (Dong et al., 2012) and changes occurring in durum wheat with a focus on allergens were analyzed (Arena et al., 2017). These authors based their methodology on two-dimensional electrophoresis (2-DE) and two-dimensional differential gel electrophoresis (2D-DIGE). However, these two approaches have several limitations such as identification of a low number of proteins. Other methods, such as isobaric tags for relative and absolute quantitation (iTRAQ)-based quantitative proteome approach (Ma et al., 2014) have also been used to analyze the proteome of wheat grain development. Recently, Victorio et al. (2018) used nanoUPLC and Ultra Definition Mass Spectrometry to identify and quantify soluble proteins in nine bread wheat cultivars of low, medium or high quality.

The nutritional status of the plant is known to significantly affect protein synthesis and accumulation in wheat. In particular, nitrogen (N) and sulfur (S) nutrition have been reported to affect several enzymes in wheat grain. For example, a glyceraldehyde-3-phosphate dehydrogenase and a serpin were both increased using a treatment combining high N and low S (Flæte et al., 2005). Modifications of AG synthesis and accumulation in the grain due to different levels of N and S likely disturbed the balance of cell functions that differentially affected carbohydrate and amino acid metabolism, and transport (Bonnot et al., 2017).

N and S supplies together influence the N to S ratio in the grain, leading to significant changes in grain SP composition at maturity (Dai et al., 2015; Bonnot et al., 2017). N supply induces changes in the rate and duration of SP accumulation during grain filling. SP concentration is increased by higher N supply and the proportion of different SP classes is modified in response to N application (Chope et al., 2014; Bonnot et al., 2017; Yu et al., 2017). S deficiency also has strong impacts on SP composition, which is associated with alterations of dough rheology (Zhao et al., 1999). S fertilization improved all major parameters of baking quality in wheat by increasing the content of S-containing amino acids (Järvan et al., 2008). S deficiency promotes the synthesis and accumulation of S-poor proteins such as ω -gliadin and HMW-GS at the expense of S-rich proteins (Zhao et al., 1999; Flæte et al., 2005; Bonnot et al., 2017).

In contrast to this, the effects of S on the synthesis and accumulation of HMW and LMW glutenins are opposite. Therefore, S deficiency decreases the total amount of polymeric proteins because LMW are the major components of glutenins. Changes were brought on by an imbalance of S-modified protein composition which is associated with alterations of dough rheology (Zhao et al., 1999). S fertilization improved all major parameters of baking quality in wheat by increasing the content of S-containing amino acids (Järvan et al., 2008).

In summary, N and S supplies together influence the N to S ratio in the grain leading to significant changes in SP composition at maturity (Bonnot et al., 2017).

In order to gain insights into the contribution of AGs to grain storage composition, we analyzed the large-scale proteome dataset for Einkorn (*Triticum monococcum* ssp. *monococcum*) described in Bonnot et al. (2017). Einkorn is an ancestral wheat whose diploid genome is a sister to bread wheat (*T. aestivum*) genome A (Marcussen et al., 2014) and for which genomic resources are available, making it a suitable species in which to explore proteome responses to N and S nutrition. We used the integrative methods implemented in the R package MixOmics (Lê Cao et al., 2016) and put forward by Duruflé et al. (2018) that take into account all quantified proteins and not only those differentially expressed as in classical omics analyses. We identified 18 important AGs. We provided a statistical validation of five of them in bread wheat by association analysis and seven of these identified AGs have already been identified as potential candidates in a network analysis based on association rule discovery (Bonnot et al., 2017). This cross-validation confirms the ability of these integrative methods to identify robust candidate entities, which could be considered for functional analysis.

MATERIALS AND METHODS

Plant Material and Grain Sampling and Processing

The einkorn wheat (*T. monococcum*) accession ERGE 35821 was cultivated in a greenhouse as described previously (Bonnot et al., 2017). Briefly, plants were grown in PVC columns filled with a 2:1 (v:v) mixture of washed perlite and river sand and arranged to form an homogenous canopy with a density of 512 plants m^{-2} . The experimental design was a randomized complete block design with four blocks and four treatments. Until anthesis, each PVC column received 167 mL $days^{-1}$ of a modified Hoagland's nutrient solution (Castle and Randall, 1987) containing 3 mM N and 0.1 mM S prepared with demineralized water. 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 to 700°C days after anthesis, four N and S treatments were applied: N0S0, nutrient solution with no N or S; N, 6 mM N with no S; S, low N (0.5 mM) and high S (2 mM); or NS, high N (6 mM) and high S (2 mM) as described by Bonnot et al. (2017) (**Supplementary Figure 1**).

Main-stem ears were tagged at the date of anthesis and only grains from the middle of the tagged ears were harvested every

100°C days (approximately every 5 days) from 100 to 1000°C days after anthesis. Thermal time was calculated as cumulative degree-days above 0°C. For each treatment and sampling date, grains from four to 10 main-stem ears were collected per replicate, depending on the grain developmental stage. Except for those used to determine grain dry mass, samples were frozen immediately in liquid N₂, and then stored at -80°C until use. Four biological replicates of whole grains were used to determine grain dry mass and SP composition, and three to identify and quantify AG proteins.

Measurement of Grain Traits

Gliadin and glutenin proteins were sequentially extracted from 100 mg of wholemeal flour milled from each replicate of grain sampled between 300 and 1000°C days after anthesis, as described by Plessis et al. (2013), Bonnot et al. (2017). Gliadin classes and glutenin subunits were separated and quantified by reverse phase high-performance liquid chromatography following the procedure described by Dai et al. (2015). Chromatograms were processed with ChemStation 10.1 software (Agilent Technologies) and the peaks corresponding to each of the four gliadin classes and the two glutenin subunits were identified following the observations of Wieser et al. (1998).

Analysis of Metabolic Proteins

Metabolic proteins were extracted from whole grains as described in Bonnot et al. (2017). Briefly, proteins were extracted for 2 h at 4°C in 10 mM sodium phosphate, 10 mM NaCl, pH 7.8, supplemented with a cocktail of plant protease inhibitors (Sigma, St. Louis, MO, United States). After centrifugation at 8,000 × 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 and afterward stored at -20°C. The precipitated AG proteins were suspended in a solubilization buffer (0.1% ZALS I, 6 M urea, 2 M thiourea, 10 mM DTT, 30 mM Tris-HCl pH 8.8, 50 mM NH₄HCO₃). Proteins were digested in-solution by trypsin and resulting peptides were analyzed by LC-MS/MS using a nanoLC Ultra system (Eksigent) and a Q-Exactive mass spectrometer (Thermo Electron).

Protein Identification and Quantification

The mass spectrometry proteomic data were first published in Bonnot et al. (2017) and have been deposited to the ProteomeXchange Consortium (Deutsch et al., 2017) via the PRIDE (Vizcano et al., 2016) partner repository with the dataset identifier PXD006058.

Proteins were identified by matching peptides using the Uniprot protein database version 2014_07 limited to the *Triticum* genus using X!Tandem¹ and X!TandemPipeline (Langella et al., 2017) as described in Bonnot et al. (2017).

Functional classification was established on the basis of gene ontology (GO) information rules provided by Uniprot (Ashburner et al., 2000). Proteins that had no

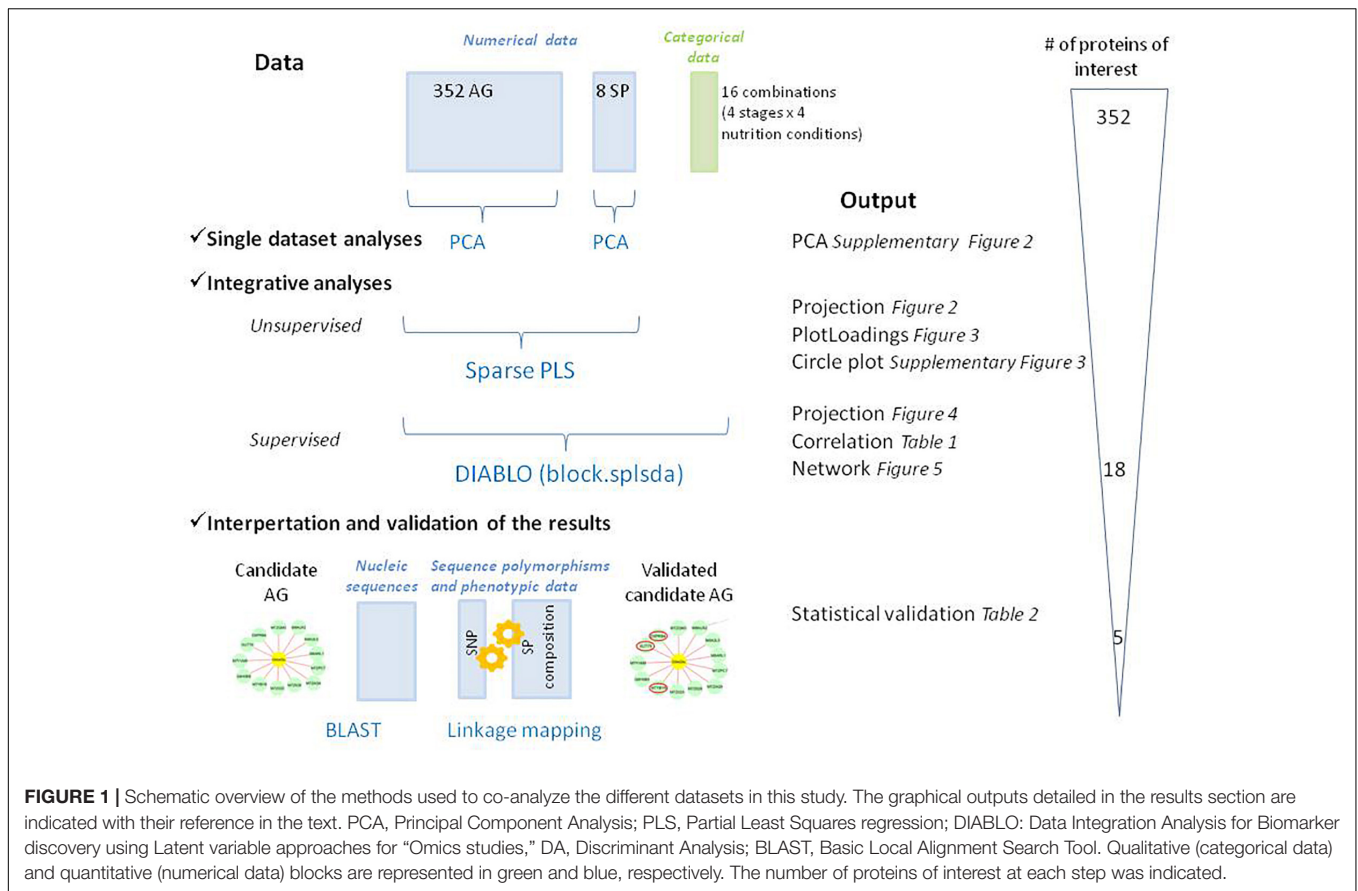
functional annotation or GO information were analyzed with Blast2GO (version 3.2, Conesa et al., 2005) in order to assign potential function.

Peptide quantification was performed by integration of extracted ion current (XIC) using MassChroQ (Valot et al., 2011). Protein relative quantities were computed by summing the XIC value of specific peptides, i.e., peptides that were not shared by different proteins. Only proteins quantified with at least two peptides were selected for subsequent analysis.

Data Integration and Discriminant Analysis

Omics data integration was performed using the R package MixOmics (version 6.3.2), which offers a whole range of multivariate methods to explore and integrate biological datasets (Lê Cao et al., 2009; Rohart et al., 2017). Our datasets consisted of two sets (the AG and the SP sets) measured on the same samples with three independent biological replicates. The AG dataset included 352 proteins present in all samples and in the 2018_12 version of Uniprot protein database. The SP dataset included the quantity per grain of total glutenin subunits and gliadins, the total quantity per grain for each glutenin and gliadin classes i.e., HMW and LMW, α-, γ-, and ω-gliadins. These two datasets (values were scaled) were explored according to the workflow presented (Figure 1). We first investigated independently each dataset by Principal components analysis (PCA). Then, both datasets were analyzed conjointly with an unsupervised method, the sparse Partial Least Squares (sPLS) regression (Lê Cao et al., 2008). The PLS algorithm searches for the largest covariance between orthogonal components, which are linear combinations of the variables from both datasets studied. Sparsing leads to the selection of variables, which are the most involved in the relationships between the biological datasets. We kept 15 variables for each dataset (i.e., all the variables for the SP set). The *plotIndiv()* function was used to project the samples onto the PCA and sPLS components. For sPLS, two plots can be obtained because a set of PLS-components is calculated on each dataset. Mathematically, it corresponds to two subspaces spanned by the variables in each dataset. A mean subspace can result from them. In this third subspace, sample coordinates were averaged from the subspaces spanned by each dataset. The *plotloading()* function was used to visualize the loading vector of the sPLS analysis and represent the contribution of the variables kept for each component. Then, we used the *block.splda()* function from DIABLO (Data Integration Analysis for Biomarker discovery using Latent variable approaches for Omics studies) (Singh et al., 2019) to identify correlated (or co-expressed) variables measured on our two datasets, which also explain the categorical variable of interest used to supervise the analysis. This latter analysis was supervised by crossing the developmental stage of the grain (four stages) and the nutrition (four nutritions) resulting in 16 combinations. Networks were built using the function *network()* from the package mixOmics to show the correlations ($r > |0.8|$) between selected features. Cytoscape software version 3.7.1 (Shannon et al., 2003) was used to visualize the network.

¹www.thegpm.org/tandem



Statistical Validation

A statistical validation of AGs involved in the network was undertaken by linkage mapping. The coding sequence of each AG in the network was retrieved from Uniprot and then blasted against the wheat pseudomolecule (IWGSC RefSeq v1.1 database²). The BLAST results were analyzed to find the coordinates of the bread wheat homologs on the pseudomolecule. Taking into account flanking sequences could be relevant because of linkage disequilibrium range in wheat and their role in gene expression. Therefore, gene regions were increased by 1,500 nucleotides upstream and downstream. Single Nucleotide polymorphisms (SNPs) located in these increased regions were extracted from the set developed by Rimbart et al. (2018). Their genotyping data were retrieved for the 196 lines from the agronomic part of the INRA worldwide core collection (Bordes et al., 2011), which was phenotyped for grain SP composition (Plessis et al., 2013). Association study was performed with the mixed model (Yu et al., 2006) implemented in the R package GWAS from rrBLUP (Endelman, 2011) comprising the kinship matrix (K) that accounts for relatedness among accessions to limit spurious associations. The Leave One Chromosome Out (LOCO) approach was used to construct different kinship matrices by testing each chromosome and leaving out the SNPs on the chromosome

being tested (Yang et al., 2014). Each off-target variants (OTV) was recoded to create two biallelic markers (a first one with the two nucleotide variants and a second one coded as presence for the nucleotide variants) or absence (for no call). Markers with a minor allele frequency < 0.05, and a missing rate > 0.1 were discarded. The significance of associations was tested with an *F*-test. Associations were judged significant at $P < 0.001$.

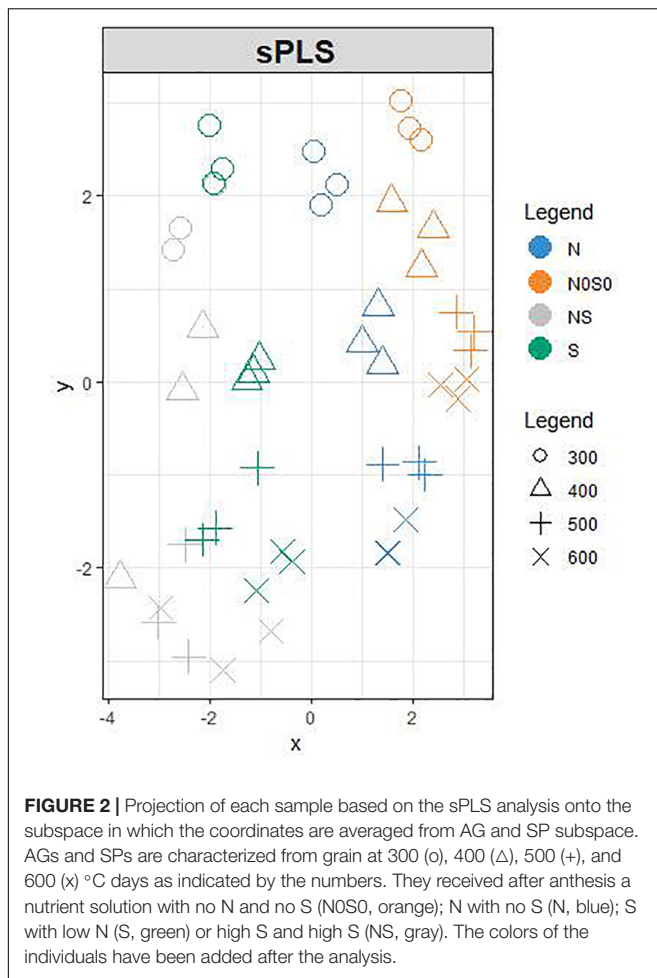
RESULTS

The AG proteome of the grain in development in response to N and S nutrition was first analyzed by Bonnot et al. (2017) to emphasize components involved in SP synthesis. Here, both dataset (AGs included 352 quantified proteins present in all samples and still present in Uniprot database, and SPs) were integrated to find candidate AGs related to seed storage synthesis.

Data Structuration

We used PCA (Supplementary Figure 2) and sPLS (Figure 2) to observe the structuration of our datasets. Individuals were projected in the subspace consisting of the AG and SP sets and then in a subspace in which coordinates are averaged from the two precedent subspaces (Figure 2). Roughly, grain developmental stages were discriminated in the AG subspace

²<https://www.wheatgenome.org/>



with the first two principal components explaining up to 54% of the total variance, while nutrition conditions were better discriminated in the subspace defined by SPs with 95% of the variability explained by the two first principal components (Supplementary Figure 2). The developmental stages and nutrition conditions are rather well discriminated in the mean subspace resulting from sPLS, which integrated both AG and SP datasets (Figure 2). In this subspace, the *x*-axis discriminated the nutrition treatments. The two opposite nutrition conditions (N0S0 and NS) were at both extremities of this axis, which could be considered as an *S*-axis. Indeed, nutrition conditions with *S* (NS and S) have negative *x*-coordinates unlike conditions lacking *S* (N and N0S0). The *y*-axis discriminated the developmental stages with 300 and 600°C days having, respectively, positive and negative coordinates. The loading weights indicated that SPs strongly contributed to the first two components (Figure 3). For the first component, all SP variables had an absolute weight >0.2, a positive weight being observed only for the gliadin to glutenin ratio. The quantity of HMW glutenin was the only SP variable with a positive contribution for the second component. This contribution was light. Six AGs (out of the 15 kept for this component) have a contribution for the first component with an absolute weight >0.2. The strongest weights (>0.5)

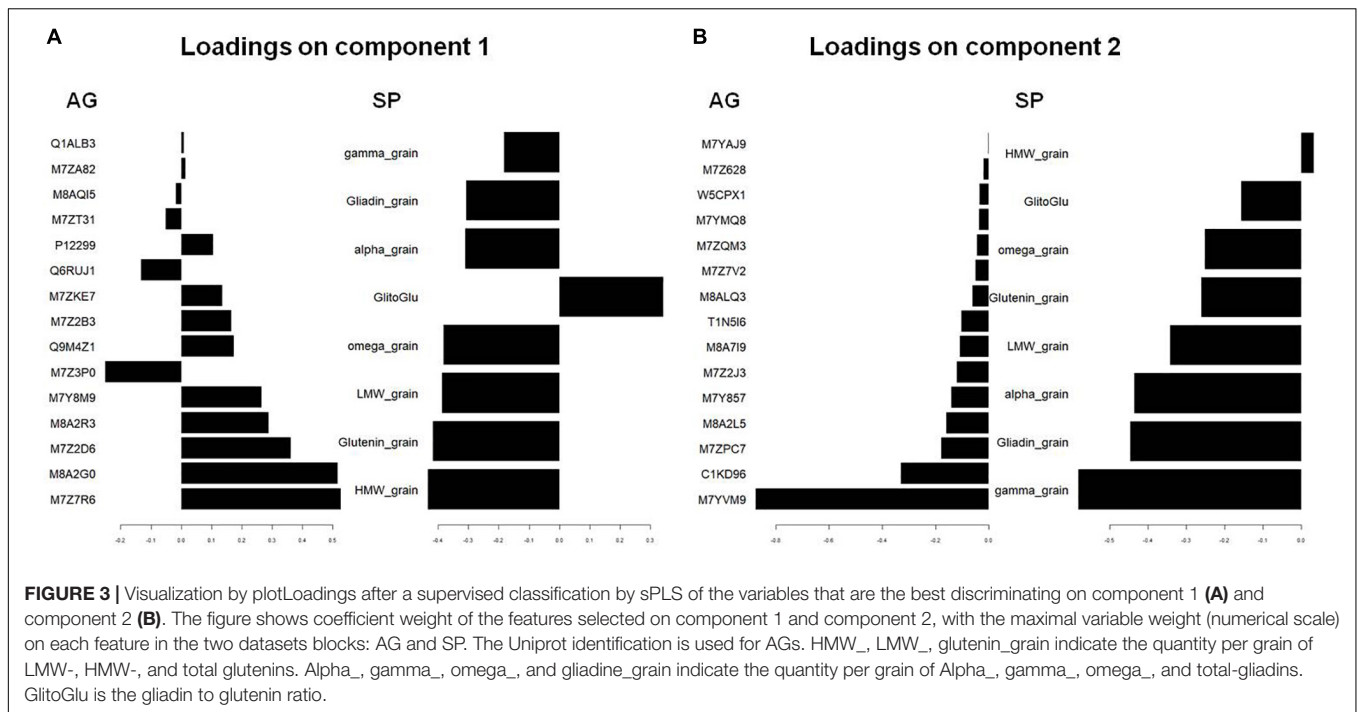
are found for M7Z7R6 and M8A2G0. M7Z3P0 was the AG with the most important negative contribution for this axis. M7YVM9 had the strongest negative contribution to the second axis (>-0.8). According to these contributions for the two first components, the *x*-axis of the circle plot (Supplementary Figure 3) is mostly explained by all variables related to SP composition, which appear to be influenced by the nutritional status. The gliadin to glutenin ratio was negatively correlated with the other SP variables. We observed that some AG variables are close in the correlation circle plot to SP variables (Supplementary Figure 3). For instance, the gliadin to glutenin ratio is close to M7ZKE7 (Uncharacterized protein/glutathione transferase domain) and M8A2R3 (Uncharacterized protein/Ricin B-like lectin EULS3-like family).

As expected, the DIABLO analysis with the *block.splda()* function improved the structuration resulting from the sPLS analysis as it is a supervised method aiming at discriminating the 4 groups (Figure 4). In the subspace of AGs, the *x*- and *y*-axis discriminated the nutrition treatments and the developmental stages, respectively (Figure 4A). The nutrition treatments were clearly discriminated in the SP subspace (Figure 4A), where the separation of the development stage was better for the *S* conditions (NS and S) than for the *N* conditions (N0S0 and N). The correlation circle plot (Figure 4B) highlighted the contribution of selected variables to each component. SPs are related to the *x*-axis with a clear opposition between the quantity per grain of glutenin subunits (HMW and LMW) and total glutenin, and the gliadin to glutenin ratio. Some AGs appeared to be highly correlated with specific SP variables. For instance, M7Z534 (Glutaredoxine-C8) and Q6RUJ1 (Glutamine synthetase) are positively correlated with the quantity per grain of HMW glutenins. Similarly, M8AUX2 (Uncharacterized protein/caleosin related domain) is positively correlated with the gliadin to glutenin ratio. Many AGs were projected as a cluster in the correlation circle plot (Figure 4B). The *x*-axis coordinate for this cluster is near 0 indicating that these AGs are not affected by nutrition status. The *y*-coordinate is strongly negative indicating a higher accumulation in the late developmental stages.

Network Analysis Reveals Albumins-Globulins Involved in Grain Storage Protein Accumulation

We analyzed the results of the two-block analysis performed with the *block.plsda* from DIABLO, supervised by the nutrition × stage combinations, where the AG and SP datasets were quantitative variables with three independent biological replicates analyzed. Fifteen AG proteins were selected on each component and connected to variables present in the SP block. On the network (Figure 5 and Table 1), positive and negative correlations > |0.8| between variables from 18 AG to 7 SP variables from datasets were simultaneously represented. Three sub-networks were identified.

The main sub-network includes 16 AGs, which were connected to two SP variables (gliadin to glutenin ratio and the quantity of HMW glutenins per grain). In this sub-network, the correlations between the gliadin to glutenin ratio and the 13



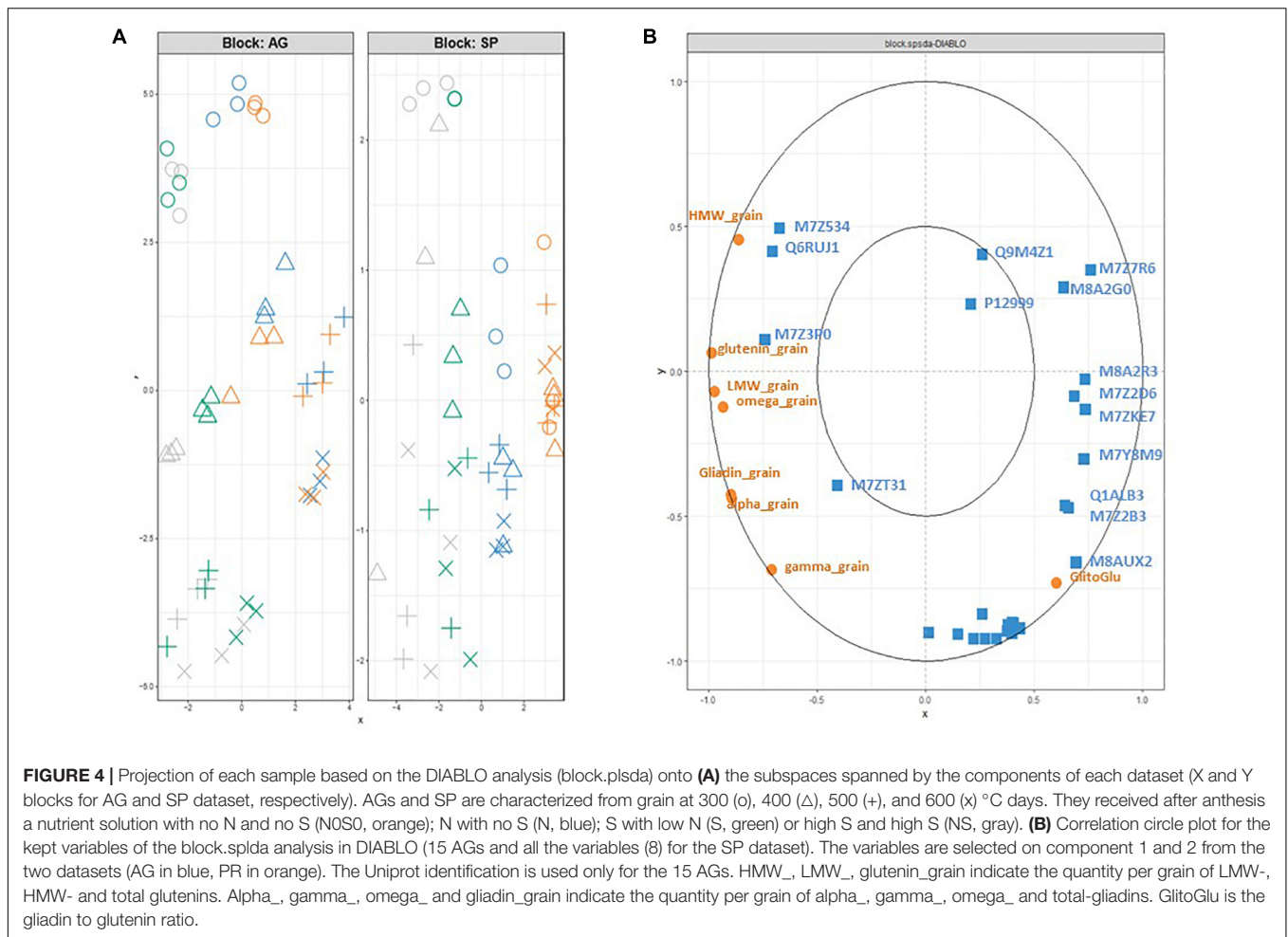
AGs linked with the SP variables were all positive (Table 1). Out of these 13 metabolic proteins, two were uncharacterized proteins (M8A2L5, M8ARL1), which nevertheless contained annotated domains. M8ARL1 has an oil body associated A1 domain and M8A2L5 classified in water stress induced proteins. The 11 remaining AGs have an annotation. M7ZQM3 is a globulin 1S. I0JTT9 is a glutamate decarboxylase known to be the rate-limiting enzyme for γ -aminobutyric acid (GABA) synthesis. M7YBY6 and M7YVM9 are two chitinase proteins related to the pathogenesis-related proteins and also involved in plant abiotic stress responses. M7ZAQ9 is an oxalate oxidase, which catalyzes the oxidation of oxalic acid to CO_2 , Ca^{2+} , and H_2O_2 , and is stored as insoluble calcium salt upon reaction with molecular oxygen (Lane et al., 1993). Q8H0B8 is a cold regulated protein, member of the phosphatidylethanolamine-binding protein (PEBP) superfamily. M7Z025 is an ankyrin repeat domain-containing protein 2. Ankyrin repeats are present in many proteins of eukaryotes, prokaryotes and some viruses and they function as protein-protein interaction domains. D0PRB4 (Peroxiredoxin) and M7Z628 (1-Cys peroxiredoxin PER1) are known to play a significant role in protection systems such as peroxidases or molecular chaperones; M7ZPC7 is, a heat shock cognate 70 kDa protein, able to act as a chaperone protein, playing major roles in the prevention of target protein aggregation and promotion of their correct folding. The last AG with an annotation was M8AUX2, a caleosin which is, like caleosin-related sequences, a lipid-associated protein found in a wide range of higher plants. M8AUX2 played a particular role in the sub-network as the AG connecting the two SP variables with a positive correlation to the gliadin to glutenin ratio and a negative one to the quantity of HMW glutenins. This latter SP variable was negatively correlated to M7Y8M9,

a translocon-associated protein subunit beta and M7Z2B3, a transmembrane 214-B domain, involved in endoplasmic reticulum-stress induced apoptosis, and positively correlated with Q6RUJ1, a glutamine synthetase.

The second sub-network concerned M7Z7R6, an uncharacterized protein, similar to some isoflavone reductase, which showed strongly negative correlations with all S-rich SPs (α -, γ -, total-gliadins, and LMW glutenins). This grouping could result from the common response to S supply of these SP classes.

The last sub-network was the smallest. It included two negatively correlated components, the ω -gliadin (S-poor SPs) and M8A2R3. M8A2R3, an uncharacterized protein, presents a domain related to ricin B-like lectin EULS3-like family. This domain is found in a group of plant proteins called lectins which bind carbohydrates.

We clearly observed that abundance of the 18 AGs from the network was more impacted by the grain developmental stage than by the N and S treatments (Supplementary Figure 4). All of them, except M7Y8M9 (Translocon-associated protein subunit beta), M8A2R3 (Uncharacterized protein/Ricin B-like lectin EULS3-like family), M7Z7R6 (Uncharacterized protein/isoflavone reductase), and Q6RUJ1 (Glutamine synthetase), presented a pattern of accumulation, with a continuous increase over development. Conversely, Q6RUJ1 (Glutamine synthetase) had an opposite accumulation pattern (decreasing from 300 to 600°C days after anthesis). The accumulation of M7Z7R6 (Uncharacterized protein/isoflavone reductase) was constant over time. Considering the response of the AGs to the different N/S supplies, we could observe that these four proteins had different profiles. M7Y8M9 (Translocon-associated protein subunit beta), M8A2R3 (Uncharacterized protein/Ricin B-like lectin EULS3-like family), contents



decreased with S supply whereas Q6RUJ1 (Glutamine synthetase) and M7Z7R6 (Uncharacterized protein/isoflavone reductase) ones increased under the two high S treatments (S and NS). However, the accumulation for the latter AG is significantly increased by high N supply (treatment N).

Association Study for Statistical Validation of the 18 AGs Highlighted by DIABLO Approach

To validate the results evidenced in our data analysis, a genetic association study was done by searching in the bread wheat genome for the orthologous and paralogous genes coding for the 18 AGs highlighted in the network. A search for genes coding the 18 AGs included in the network was carried out in bread wheat genome (**Supplementary Table 1**). Eighty-eight loci orthologous or paralogous were detected. All chromosomal groups were represented except the group 6. Using coordinates of each gene, we extracted 181 SNP markers and 11 OTV from the BreedWheat set (**Supplementary Table 1**). After filtering, 126 markers in 34 genes out of 81 (42%) have been used for analysis. No marker was available in the gene coding for M8A2R3, M7ZQM3 and M7ZAQ9.

Significant associations ($P < 0.001$) were found in a least one of the homologous gene coding for five AGs (D0PRB4, I0JTT9, M7YBY6, M7Z2B3, and M7Z7R6) (**Table 2**). D0PRB4 (Peroxiredoxin) was associated with the quantity of ω -gliadin per grain; I0JTT9 (Glutamate decarboxylase) to the quantity of γ -gliadin per grain; M7YBY6 (Basic endochitinase C) to quantities per grain of LMW glutenins and gliadins (especially the α -gliadin); M7Z2B3 (Uncharacterized protein/transmembrane 214-B) to the quantity of HMW glutenins per grain and thus to the HMW to LMW glutenin ratio; M7Z7R6 (Uncharacterized protein/isoflavone reductase) to the quantity of ω -gliadin per grain. Therefore, linkage mapping confirmed the role of some AGs in the synthesis and accumulation of SPs. It is worth noting that associations found for M7Z2B3 (Uncharacterized protein/transmembrane 214-B) perfectly fitted the network produced by DIABLO analysis (**Figure 5**).

DISCUSSION

From a dataset of 352 AGs, we were able to take into account the effect of the combination of stage of development and N and S supply to identify the most discriminating proteins

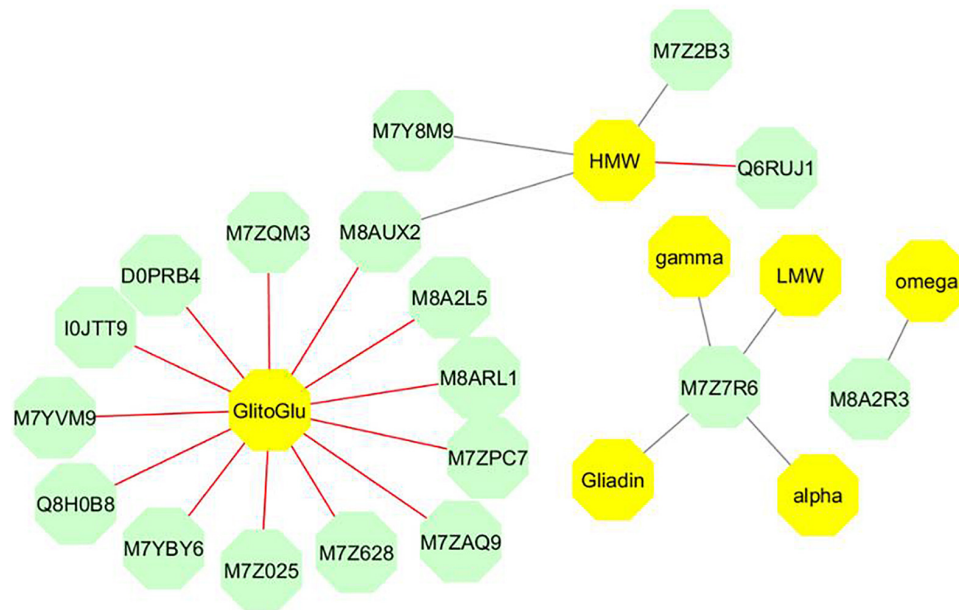


FIGURE 5 | Network produced by the block.plsda in DIABLO. Each color represents a type of variable (green and bright yellow for AGs and SPs, respectively). The correlation cut off applied was 0.8. The optimal components were considered in the network, which is bipartite i.e., only links between two variables of different types are drawn. The color of edges indicates the sign of the correlation: red and gray correspond to positive and negative correlation, respectively (Table 1). The Uniprot identification is used for AGs. HMW and LMW indicate the quantity per grain of HMW- and LMW- glutenins. Alpha, gamma, omega and Gliadin are the quantity per grain of α -, γ -, ω , and total-gliadins, respectively. GlitoGlu is the gliadin to glutenin ratio.

associated with SP composition using sparse methods. Five of these proteins were statistically validated by genetic association analysis. These results demonstrate the capacity of these new methods to identify candidates for further investigations, especially functional validation. Due to the clear structuration of the data that was found to be related to developmental stage and nutrition, quantitative data were driven using these two factors. Following the identification of the most discriminating AGs, a network was established to focus on the candidate AGs highly correlated to SPs.

Input of the Integrative Approach

Data-mining methods are diverse. For this study, we used a multivariate projection methodology developed in DIABLO to produce a network, which includes 18 AGs and 7 variables related to seed SPs composition. Roughly, these 18 AGs are involved in N metabolism, stress sensing, storage metabolism and, protein processing or folding (Figure 6). Other methods for data integration are available such as methods based on the discovery of association rules between attributes (Agiar et al., 2007; Vincent et al., 2015). This method has been used to find associations between proteins and to compare their behavior between treatments by Bonnot et al. (2017). The network obtained by these authors included several AGs, including 7 AGs (M7YBY6, M7YVM9, M7Z2B3, M8AUX2, M8A2R3, M7Y8M9, and Q6RUJ1) that were also highlighted by the multivariate projection methodology (Figure 6). The statistical validation carried out using genetic association analysis confirmed the putative role of 5 candidate AGs out of the 18,

D0PRB4, I0JTT9, M7YBY6, M7Z2B3, M7Z7R6, in the synthesis of SPs (Figure 6). M7YBY6, annotated as an endochitinase, and M7Z2B3 (involved in the reticulum stress apoptosis) are identified by both integrative methods and significantly associated with SPs. Altogether, the integration of AG and SP datasets provided firm and consistent results on the relationship between some AGs and SPs. A few central components have been highlighted. Therefore, biological interpretation of the network could be helpful to improve understanding of the SP synthesis.

Biological Interpretation

Nutrients, stress, hormones, and growth factors all affect protein biosynthesis through complex signaling pathways. Although the synthesis of AGs is less affected by N than that of SPs (Wieser and Seilmeier, 1998; Martre et al., 2003; Plessis et al., 2013), N and S nutrition has been reported to affect several enzymes in wheat grain (Flæte et al., 2005). Recently, Bonnot et al. (2017) have shown that the grain proteome was impacted by the N/S balance. Our results revealed dynamic enzyme activities related to the grain development (four stages during the early phases of endosperm cell division and cell differentiation) and the N/S supply (four nutrition supplies). The early phases of grain development determine the potential final size of the grain (Xie et al., 2015) and are particularly sensitive to environmental conditions. The multivariate projection method used in this work revealed variables from each dataset (AGs and SPs) mainly involved in discrimination according to the combination of N and S supply and grain developmental stage (Figure 4B). Eighteen AGs and seven SP variables important

TABLE 1 | Proteins highlighted in the network.

| Uniprot ID | Protein name ^a | LMW_glutenin | HMW_glutenin | α _gliadin | γ _gliadin | ω _gliadin | Gliadin_grain | GliToGlu |
|------------|--|--------------|--------------|-------------------|-------------------|-------------------|---------------|----------|
| D0PRB4 | Peroxiredoxin | | | | | | | 0.91 |
| I0JTT9 | Glutamate decarboxylase | | | | | | | 0.92 |
| M7Y8M9 | Translocon-associated protein subunit beta | | -0.83 | | | | | |
| M7YBY6 | Basic endochitinase C | | | | | | | 0.83 |
| M7YVM9 | Chitinase 1 | | | | | | | 0.82 |
| M7Z025 | Ankyrin repeat domain-containing protein 2 | | | | | | | 0.85 |
| M7Z2B3 | Uncharacterized protein/transmembrane 214-B* | | -0.80 | | | | | |
| M7Z628 | 1-Cys peroxiredoxin PER1 | | | | | | | 0.86 |
| M7Z7R6 | Uncharacterized protein/isoflavone reductase* | -0.82 | | -0.83 | -0.82 | | -0.82 | |
| M7ZAQ9 | Oxalate oxidase 2 | | | | | | | 0.88 |
| M7ZPC7 | Heat shock cognate 70 kDa protein | | | | | | | 0.90 |
| M7ZQM3 | Globulin-1 S allele | | | | | | | 0.91 |
| M8A2L5 | Uncharacterized protein/water stress induced protein * | | | | | | | 0.91 |
| M8A2R3 | Uncharacterized protein/Ricin B-like lectin EULS3-like family* | | | | | -0.81 | | |
| M8ARL1 | Uncharacterized protein/Oil body-associated 1A* | | | | | | | 0.90 |
| M8AUX2 | Uncharacterized protein/caleosin 2* | | -0.91 | | | | | 0.88 |
| Q6RUJ1 | Glutamine synthetase | | 0.85 | | | | | |
| Q8H0B8 | Cold regulated protein | | | | | | | 0.89 |

Accessions and protein names given by Uniprot were indicated. Only the positive and negative correlations values ($>|0.8|$) obtained between AGs and SP with the DIABLO approach were given. Storage proteins expressed by grain (mg N grain⁻¹), GliToGlu, gliadin to glutenin ratio. ^a*Indicates function assigned by BLAST for uncharacterized proteins.

TABLE 2 | Association analysis. Proteins from the network associated to storage protein composition.

| Uniprot ID | Chr | HMWtoLMW | HMW | LMW | Gliadin | α _gliadin | γ _gliadin | ω _gliadin |
|------------|-----|----------|-----|-----|---------|-------------------|-------------------|-------------------|
| D0PRB4 | 2B | | | | | | | X |
| I0JTT9 | 3A | | | | | | x | |
| M7YBY6 | 7A | | | x | | | | |
| M7YBY6 | 7B | | | | x | x | | |
| M7Z2B3 | 3A | x | x | | | | | |
| M7Z7R6 | 1A | | | | | | | X |

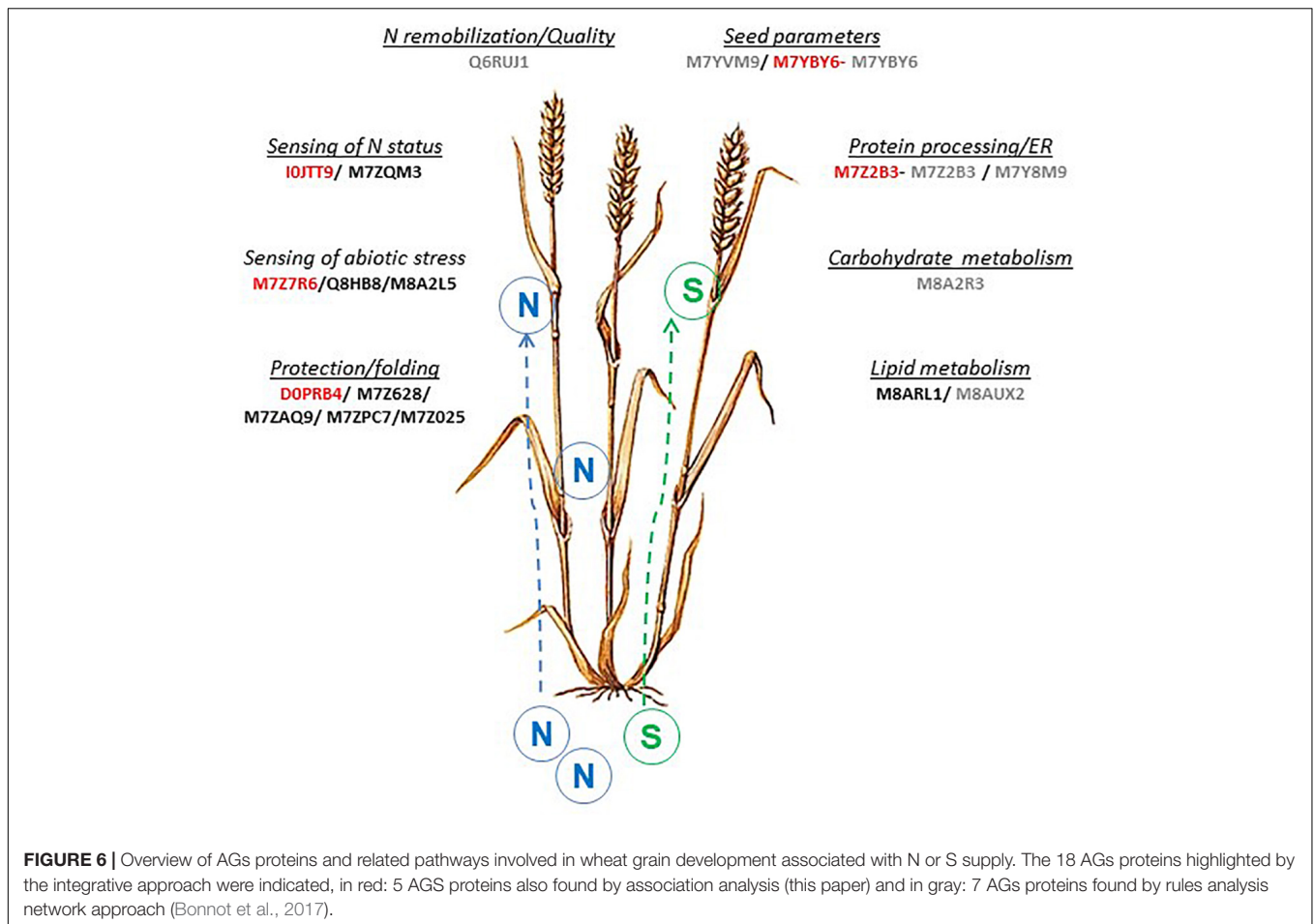
Uniprot ID, accession Uniprot; Chr, chromosome; HMWtoLMW, HMW glutenins to LMW glutenins ratio; HMW, LMW, Gliadin, quantity of storage proteins by grain (mg N grain⁻¹); α _, γ _, ω _gliadin, percentage of alpha-, gamma-, omega-gliadin in total gliadins.

in discrimination were highly correlated as shown by the network. Based on the principle of “guilt per association,” these AGs could influence the synthesis of SPs in relation to developmental stage and nutrition as confirmed for a few of them by linkage mapping. These proteins are involved in sensing processes, N assimilation, protein protection or folding, and storage metabolism.

Sensing of Abiotic Stress

Therefore, several AGs linked to SPs have a function related to S or N starvations [M7Z7R6 (Uncharacterized protein/isoflavone reductase), Q6RUJ1 (Glutamine synthetase), I0JTT9 (Glutamate decarboxylase), and M7ZQM3 (Globulin-1 S allele)].

M7Z7R6 has similarities with an isoflavone reductase. In maize (*Zea mays*), a gene highly homologous to isoflavone reductase has been identified; it encodes an NADPH binding enzyme and is activated in response to S starvation (Petrucco et al., 1996). The glutathione-dependent regulation indicated that maize isoflavone reductase like genes may play a crucial role in the establishment of a thiol-independent response to oxidative stress under glutathione shortage conditions. The identification of genes activated in S starvation may thus provide insights not only into the mechanisms of adaptation to nutrient limitation but also into the response to stress resulting from glutathione depletion. In rice, an isoflavone reductase-like gene has been shown to be involved in homeostasis of reactive oxygen species (Kim et al., 2010). In wheat, detailed physiological



dissection of this gene is underway to identify the underlying mechanism conferring heat tolerance (Singh et al., 2018). The location of M7Z7R6 in the correlation circle (**Figure 4B**) at the right of the x -axis confirmed the role of this metabolic protein in response to S deficiency. This AG is over-expressed when S is lacking whatever is the developmental stage of the grain (**Supplementary Figure 4**). As expected, M7Z7R6 is negatively correlated with S rich SPs, α -, γ - and total-gliadins (**Figures 4, 5**).

Glutamine synthetase (Q6RUJ1) plays a major role in the assimilation of ammonia during remobilization of N to the grain (Bernard et al., 2008; Zheng et al., 2018). Q6RUJ1 projection (**Figure 4**) indicates that this enzyme is more expressed in the early development stages and in nutrition conditions with high S (NS and S). The abundance of this protein was highly correlated with that of HMW-GS.

Glutamate decarboxylase is a key enzyme of the GABA pathway (Miflin and Habash, 2002). GABA has been associated with various physiological responses (Carillo, 2018), including N starvation (Recht et al., 2014) and sensing of N status (Lancien and Roberts, 2006). In plants, GABA increase is a common response to a restriction of glutamine synthesis, reduction of protein synthesis and an increase in protein degradation (Bouché and Fromm, 2004), which happen concomitantly.

GABA is involved in 14-3-3-genes regulation (Lancien and Roberts, 2006). Shin et al. (2011) have reported the role of 14-3-3 proteins in coordinating carbon (C), S, and N metabolism through the regulation of several key enzymes including glutamine synthetase (Q6RUJ1) and glutamate decarboxylase (IOJTT9).

Globulin 1S allele (M7ZQM3) was correlated to gliadin to glutenin ratio. The gene coding for this protein was identified by Zheng et al. (2018) in response to N application.

In addition, we detected two other proteins related to abiotic stresses (M7YBY6 and M7YVM9). This is not surprising as many proteins with stress and defense-related ontologies have been reported to be upregulated during grain development (Nadaud et al., 2010; Capron et al., 2012; Ma et al., 2014; Brinton and Uauy, 2018).

Adaptative Response of the Grain to N and S Supply

During grain filling, two major processes, storage synthesis and accumulation take place. They concern starch and SPs. Grain SPs were synthesized on the rough endoplasmic reticulum (ER) and then passed into the lumen, where cleavage of signal peptides, formation of intra- and inter-chain disulphide bonds, folding,

assembly and aggregation of polypeptides happen (She et al., 2011). These processes are assisted by molecular chaperons (Bertolotti et al., 2000) and chaperon proteins in the ER lumen are crucial for the biogenesis of SP bodies (She et al., 2011).

During the processing, chaperone proteins like M7ZPC7 (Heat shock cognate 70 kDa protein) transiently bind many nascent secretory proteins and probably prevents non-specific aggregation to facilitate their folding and assembly (Muench et al., 1997). Saito et al. (2009) showed that binding proteins were involved in rice prolamin deposition in ER despite the absence of the retention signal.

Peroxisredoxins (D0PRB4 and M7Z628) are thiol peroxidases with multiple functions in the antioxidant defense and redox signaling network of the cell. They could also act as chaperons (Chae et al., 2012; Lee S.L. et al., 2018).

In eukaryotic cells, one-third of all proteins must be transported across or inserted into the endoplasmic reticulum (ER) membrane by the ER protein translocon. The translocon-associated protein (TRAP) complex is an integral component of the translocon (Pfeffer et al., 2016). The protein M7Y8M9, a subunit of the protein complex located in the ER membrane, could act in this process.

Carbohydrate and lipid metabolism could be modified by both development and nutrition status as previously described (Cao et al., 2016; Bonnot et al., 2017; Yang et al., 2017). M8ARL1 and M8AUX2 (caleosin) were related to lipid metabolism. In *Arabidopsis*, seed-specific caleosins are viewed as oil-body associated proteins that possess Ca (2+)-dependent peroxygenase activity and are involved in processes of lipid degradation (Poxleitner et al., 2006). In *Arabidopsis* some seed-specific caleosins had peroxygenase activity (Partridge and Murphy, 2009).

Lee K.H. et al. (2018) showed that grain width 2 (GW2) strongly interacted with chitinase 14 (CHT14). GW2 homolog has been identified in wheat (*T. aestivum*; TaGW2). Like in rice, this protein is involved in the regulation of grain weight and width at maturity (Su et al., 2011). M7YBY6 and M7YVM9, which have a chitinase activity, could influence these parameters and indirectly the storage capacity of the grain.

The Link With the End-Used Quality of the Flour

If AGs are rich in S-containing amino acids (Zhao et al., 1999) and could influence directly the rheological properties of wheat flour (Hill et al., 2008), they also have an indirect role by influencing the SP content and composition, which are the main determinants of flour end-used (Zheng et al., 2018). As seen above, all the AGs included in the network obtained in this work could be biologically related to SP synthesis. They are likely indirect components of wheat end-used quality.

This is probably the case for Q6RUJ1 (glutamine synthetase) as reported by Gao et al. (2009), who showed that the abundance of this enzyme is higher at the early period of grain development in a good bread making quality cultivar than in a poor one. This AG could influence the flour quality by two ways. First, a high expression level of this enzyme corresponds to a high

quantity per grain of HMW glutenins as indicated by the strong positive correlation between these two variables. HMW glutenins constitute the backbone of the gluten network. Consequently, they are a major determinant of the rheological properties of dough and influence its mixing properties (Shewry et al., 2003). Secondly, the high expression of glutamine synthetase at early developmental stage of the grain was reported to facilitate the folding of gluten proteins producing more regular glutenin polymers, which might improve the gluten quality (Mifflin and Habash, 2002).

The gliadin to glutenin ratio is related to a measure of molecular weight distribution. At constant protein content, decreases in the gliadin to glutenin ratio were associated with several dough parameters with, for examples, increases in dough mixing time, maximum resistance to extension, and loaf volume and with decreases in extensibility (Uthayakumaran et al., 1999). Therefore, its optimal value depends on the process: low values are required for bread making while high values are advantageous for pastries (Marchetti et al., 2012; Barak et al., 2014). Here, we found many AGs able to help manage this ratio depending on the nutrition and developmental stage. It could be increased by a higher expression of M8AUX2 (Uncharacterized protein/caleosin 2), which concomitantly led to a decrease of HMW glutenins.

CONCLUSION

In this study, we used new tools based on multivariate statistical approaches to unravel the relationship between AGs and SPs known to influence the end-used quality of wheat flour. The results highlighted the potential role of 18 AGs in SP synthesis, which are consistent with their annotation. In addition, 12 AGs (67%) of this set have been confirmed as important components for SP synthesis by an integrative methodology based on rules or linkage mapping. These results demonstrated the efficiency of the approach used especially since the statistical validation had been limited by the lack of markers in some sequences. Therefore, the number of AGs statistically validated might be underestimated. Two AGs (M7YBY6, an endochitinase, and M7Z2B3, involved in the reticulum stress apoptosis) were identified by all the strategies. The latter is overexpressed in S-deficient conditions. This enzyme contributes to increase the gliadin to glutenin ratio by decreasing the HMW glutenin quantity per grain. The down-accumulation of this enzyme might increase HMW glutenin quantity and then the baking quality when the level of fertilization is low. However, the role of this AG must be functionally validated.

Thus, we have shown the interest of multivariate statistical approaches to extract the most significant variables from large datasets in response to abiotic constraints. Interestingly, these methods are able to consider the whole dataset instead of using only differentially expressed variables because the hypotheses rest on similar expression patterns across a set of samples indicating functional relationships. This has an important biological meaning because variables (AGs), which are not differentially expressed during grain development or in response to N or S supply, may be important actors due to interactions with

other proteins (which may be differentially expressed). These results also highlight the consistency and complementarity of the approaches and the need to take into account the proteins obtained by each of them to get a comprehensive view. To know if changes observed could be found also at a transcriptional level, transcriptomic approaches would be necessary. However, as RNA and protein abundances are weakly correlated in wheat (about 32%, Tasleem-Tahir, 2012), we would not necessarily obtain the same response.

To conclude, the innovative method used here was shown to provide an operational framework to biologists, who can then follow only a few candidate entities for functional validation. Therefore, further work has to be initiated to see how the targeted AGs influence the SP synthesis and the quality parameters in relation to N and S fertilizer applications. Their implementation in future breeding programs has also to be investigated.

DATA AVAILABILITY

The datasets generated for this study can be found in the ProteomeXchange Consortium, PXD006058.

AUTHOR CONTRIBUTIONS

PM, EB, and CR conceived and coordinated the study. EB and CR carried out the statistical analysis. SD helped for the modeling and the optimization methods. DA collected the phenotypic data. MD, TB, EB, and MZ carried out the proteomic analysis. EB, TB,

and CR wrote the manuscript. All authors edited and approved the final version of the manuscript.

FUNDING

This work was supported by a Ph.D. grant from the French Ministry of Higher Education and Research to TB and 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), and the European Union's Seventh Framework Programme (FP7/2007–2013) under the Grant Agreement Number FP7-613556.

ACKNOWLEDGMENTS

We thank R. Blanc for his help in the greenhouse, A. Faye for organizing the plant culture and planning grain sampling, and J. Boudet, M. Merlino, I. Nadaud, and S. Perrochon for their help in collecting the grains (UMR1095).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00832/full#supplementary-material>

REFERENCES

- Agier, M., Petit, J. M., and Suzuki, E. (2007). Unifying framework for rule semantics: application to gene expression data. *Fundam. Inform.* 78, 543–559.
- Arena, S., D'Ambrosio, C., Vitale, M., Mazzeo, F., Mamone, G., Di Stasio, L., et al. (2017). Differential representation of albumins and globulins during grain development in durum wheat and its possible functional consequences. *J. Proteomics* 162, 86–98. doi: 10.1016/j.jprot.2017.05.004
- 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.
- Barak, S., Mudgil, D., and Khatkar, B. S. (2014). Influence of gliadin and glutenin fractions on rheological, pasting, and textural properties of dough. *Int. J. Food Prop.* 17, 1428–1438. doi: 10.1080/10942912.2012.717154
- Bernard, S. M., Møller, A. L., Dionisio, G., Kichey, T., Jahn, T. P., and Dubois, F. (2008). Gene expression, cellular localisation and function of glutamine synthetase isozymes in wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* 67, 89–105. doi: 10.1007/s11103-008-9303-y
- Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P., and Ron, D. (2000). Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat. Cell Biol.* 2, 326–332. doi: 10.1038/35014014
- Bonnot, T., Bancel, E., Alvarez, D., Davanture, M., Boudet, J., Pailloux, M., et al. (2017). Grain subproteome responses to nitrogen and sulfur supply in diploid wheat *Triticum monococcum* ssp. *monococcum*. *Plant J.* 91, 894–910. doi: 10.1111/tpj.13615
- Bordes, J., Ravel, C., Le Gouis, J., Lapiere, A., Charmet, G., and Balfourier, F. (2011). Use of a global wheat core collection for association analysis of flour and dough quality traits. *J. Cereal Sci.* 54, 137–147. doi: 10.1016/j.jcs.2011.03.004
- Bouché, N., and Fromm, H. (2004). GABA in plants: just a metabolite? *Trends Plant Sci.* 9, 110–115. doi: 10.1016/j.tplants.2004.01.006
- Brinton, J., and Uauy, C. (2018). A reductionist approach to dissecting grain weight and yield in wheat. *J. Integr. Plant Biol.* 61, 337–358. doi: 10.1111/jipb.12741
- Cao, H., He, M., Zhu, C., Yuan, L., Dong, L., Bian, Y., et al. (2016). Distinct metabolic changes between wheat embryo and endosperm during grain development revealed by 2D-DIGE-based integrative proteome analysis. *Proteomics* 16, 1515–1536. doi: 10.1002/pmic.201500371
- Capron, D., Mouzeyar, S., Boulaflous, A., Girousse, C., Rustenholz, C., Laugier, C., et al. (2012). Transcriptional profile analysis of E3 ligase and hormone-related genes expressed during wheat grain development. *BMC Plant Biol.* 12:35. doi: 10.1186/1471-2229-12-35
- Carillo, P. (2018). GABA shunt in durum wheat. *Front. Plant Sci.* 9:100. doi: 10.3389/fpls.2018.00100
- 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. doi: 10.1093/jxb/ern218
- Chae, H. Z., Oubrahim, H., Park, J. W., Rhee, S. G., and Chock, P. B. (2012). Protein glutathionylation in the regulation of peroxiredoxins: a family of thiol-specific peroxidases that function as antioxidants, molecular chaperones, and signal modulators. *Antioxid. Redox Signal.* 16, 506–523. doi: 10.1089/ars.2011.4260
- Chope, G. A., Wan, Y., Penson, S. P., Bhandari, D. G., Powers, S. J., Shewry, P. R., et al. (2014). Effects of genotype, season, and nitrogen nutrition on gene expression and protein accumulation in wheat grain. *J. Agric. Food Chem.* 62, 4399–4407. doi: 10.1021/jf500625c
- 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. doi: 10.1093/bioinformatics/bti610
- Dai, Z., Plessis, A., Vincent, J., Duchateau, N., Besson, A., Dardevet, M., et al. (2015). Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. *Plant J.* 83, 326–343. doi: 10.1111/tpj.12881

- Debiton, C., Merlino, M., Chambon, C., Bancel, E., Decourteix, M., Planchot, V., et al. (2011). Analyses of albumins, globulins and amphiphilic proteins by proteomic approach give new insights on waxy wheat starch metabolism. *J. Cereal Sci.* 53, 160–169. doi: 10.1016/j.jcs.2010.11.001
- Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., and Ternent, T. (2017). The proteomexchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* 45, D1100–D1106. doi: 10.1093/nar/gkw936
- Dong, K., Ge, P., Ma, C., Wang, K., Yan, X., Gao, L., et al. (2012). Albumin and globulin dynamics during grain development of elite Chinese wheat cultivar Xiaoyan 6. *J. Cereal Sci.* 56, 615–622. doi: 10.1016/j.jcs.2012.08.016
- Durulé, H., Selman, M., Ranocha, P., Jamet, E., Dunand, C., and Déjean, S. (2018). A powerful framework for an integrative study with heterogeneous omics data: from univariate statistics to multi-block analysis. *bioRxiv* 357921.
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Flæte, N. E. S., Hollung, K., Ruud, L., Sogn, T., Færgestad, E. M., Skarpeid, H. J., et al. (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. doi: 10.1016/j.jcs.2005.01.003
- Gao, L. Y., Wang, A. L., Li, X. H., Dong, K., Wang, K., Appels, R., et al. (2009). Wheat quality related differential expressions of albumins and globulins revealed by two-dimensional difference gel electrophoresis (2-D DIGE). *J. Proteomics* 73, 279–296. doi: 10.1016/j.jprot.2009.09.014
- Hill, K., Horváth-Szancics, 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. doi: 10.1016/j.colsurfa.2007.06.047
- Järvan, M., Edesi, L., Adamson, A., Lukme, L., and Akk, A. (2008). The effect of sulphur fertilization on yield, quality of protein and baking properties of winter wheat. *Agron. Res.* 6, 459–469.
- Kim, S. G., Kim, S. T., Wang, Y., Kim, S. K., Lee, C. H., Kim, K. K., et al. (2010). Overexpression of rice isoflavone reductase-like gene (OsIRL) confers tolerance to reactive oxygen species. *Physiol. Plant* 138, 1–9. doi: 10.1111/j.1399-3054.2009.01290.x
- Lancien, M., and Roberts, M. R. (2006). Regulation of Arabidopsis thaliana14-3-3 gene expression by γ -aminobutyric acid. *Plant Cell Environ.* 29, 1430–1436. doi: 10.1111/j.1365-3040.2006.01526.x
- Lane, B. G., Dunwell, J. M., Ray, J. A., Schmitt, M. R., and Cuming, A. C. (1993). Germin, a protein marker of early plant development, is an oxalate oxidase. *J. Biol. Chem.* 268, 12239–12242.
- Langella, O., Valot, B., Balliau, T., Blein-Nicolas, M., Bonhomme, L., and Zivy, M. (2017). X!TandemPipeline: a tool to manage sequence redundancy for protein inference and phosphosite identification. *J. Proteome Res.* 2, 494–503. doi: 10.1021/acs.jproteome.6b00632
- Lê Cao, K. A., Martin, P. G., Robert-Granié, C., and Besse, P. (2009). Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics* 10:34. doi: 10.1186/1471-2105-10-34
- Lê Cao, K. A., Rohart, F., Gonzalez, I., Dejean, S., Gautier, B., Bartolo, F., et al. (2016). *mixOmics: Omics Data Integration Project. R package version 6.1.1*. Available at: www.mixomics.org
- Lê Cao, K. A., Rossouw, D., Robert-Granié, C., and Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. Biol.* 7:Article35. doi: 10.2202/1544-6115.1390
- Lee, S. L., Kang, C. H., Park, J. H., and Lee, S. Y. (2018). Physiological significance of plant peroxiredoxins and the structure-related and multifunctional biochemistry of peroxiredoxin 1. *Antioxid. Redox Signal.* 28, 625–639. doi: 10.1089/ars.2017.7400
- Lee, K. H., Park, S. W., Kim, Y. J., Koo, Y. J., Song, J. T., and Seo, H. S. (2018). Grain width 2 (GW2) and its interacting proteins regulate seed development in rice (*Oryza sativa* L.). *Bot. Stud.* 59, 23–30. doi: 10.1186/s40529-018-0240-z
- Ma, C., Zhou, J., Chen, G., Bian, Y., Lv, D., Li, X., et al. (2014). iTRAQ-based quantitative proteome and phosphoprotein characterization reveals the central metabolism changes involved in wheat grain development. *BMC Genomics* 15:1029. doi: 10.1186/1471-2164-15-1029
- Marchetti, L., Miguel, C., Leda, C., and Cristina, F. (2012). Effect of gluteins of different quality on dough characteristics and breadmaking performance. *Food Sci. Technol.* 46, 224–31. doi: 10.1016/j.lwt.2011.10.002
- Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., International Wheat Genome Sequencing Consortium et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345, 1250092–1250092. doi: 10.1126/science.1250092
- Martre, P., Porter, J. R., Jamieson, P. D., and Tribou, 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. doi: 10.1104/pp.103.030585
- Merlino, M., Leroy, P., Chambon, C., and Branlard, G. (2009). Mapping and proteomic analysis of albumin and globulin proteins in hexaploid wheat kernels (*Triticum aestivum* L.). *Theor. Appl. Genet.* 118, 1321–1337. doi: 10.1007/s00122-009-0983-988
- 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. doi: 10.1093/jexbot/53.370.979
- Muench, D. G., Wu, Y., Zhang, Y., Li, X., Boston, R. S., and Okita, T. W. (1997). Molecular cloning, expression and subcellular localization of a biphomolog from rice endosperm tissue. *Plant Cell Physiol.* 38, 404–412. doi: 10.1093/oxfordjournals.pcp.a029183
- Nadaud, I., Grousse, C., Debiton, C., Chambon, C., Bouzidi, M. F., Martre, P., et al. (2010). Proteomic and morphological analysis of early stages of wheat grain development. *Proteomics* 10, 2901–2910. doi: 10.1002/pmic.200900792
- Partridge, M., and Murphy, D. J. (2009). Roles of a membrane-bound caleosin and putative peroxygenase in biotic and abiotic stress responses in Arabidopsis. *Plant Physiol. Biochem.* 47, 796–806. doi: 10.1016/j.plaphy.2009.04.005
- Petrucchio, S., Bolchi, A., Foroni, C., Percudani, R., Rossi, G. L., and Ottonello, S. (1996). A maize gene encoding a NADPH binding enzyme highly homologous to isoflavone reductases is activated in response to sulfur starvation. *Plant Cell* 8, 69–80. doi: 10.1105/tpc.8.1.69
- Pfeffer, S., Dudek, J., Schaffer, M., Ng, B. G., Albert, S., Plitzko, J. M., et al. (2016). Dissecting the molecular organization of the translocon-associated protein complex. *Nat. Commun.* 8:14516. doi: 10.1038/ncomms14516
- 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. doi: 10.1093/jxb/ert188
- Poxleitner, M., Rogers, S. W., Lacey Samuels, A., Browse, J., and Rogers, J. C. (2006). A role for caleosin in degradation of oil-body storage lipid during seed germination. *Plant J.* 47, 917–933. doi: 10.1111/j.1365-313x.2006.02845.x
- Recht, L., Töpfer, N., Batushansky, A., Sikron, N., Gibon, Y., Fait, A., et al. (2014). Metabolite profiling and integrative modeling reveal metabolic constraints for carbon partitioning under nitrogen starvation in the green alga *Haematococcus pluvialis*. *J. Biol. Chem.* 289, 30387–30403. doi: 10.1074/jbc.M114.555144
- Rimbert, H., Darrier, B., Navarro, J., Kitt, J., Choulet, F., Leveugle, M., et al. (2018). High throughput SNP discovery and genotyping in hexaploid wheat. *PLoS One* 13:e0186329. doi: 10.1371/journal.pone.0186329
- Rohart, F., Gautier, B., Singh, A., and Lê Cao, K. A. (2017). mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13:e1005752. doi: 10.1371/journal.pcbi.1005752
- Saito, Y., Kishida, K., Takata, K., Takahashi, H., Shimada, T., Tanaka, K., et al. (2009). A green fluorescent protein fused to rice prolamin forms protein body-like structures in transgenic rice. *J. Exp. Bot.* 60, 615–627. doi: 10.1093/jxb/ern311
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- She, M., Ye, X., Yan, Y., Howit, C. M., Belgard, M., and Ma, W. (2011). Gene networks in the synthesis and deposition of protein polymers during grain development of wheat. *Funct. Integr. Genomics* 11, 23–35. doi: 10.1007/s10142-010-0196-x
- 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. doi: 10.1093/jexbot/53.370.947
- Shewry, P. R., Halford, N. G., Tatham, A. S., Popineau, Y., Lafandra, D., and Belton, P. S. (2003). The high molecular weight subunits of wheat glutenin and

- their role in determining wheat processing properties. *Adv. Food Nutr. Res.* 45, 219–302. doi: 10.1016/s1043-4526(03)45006-7
- Shin, R., Jez, J. M., Basra, A., Zhang, B., and Schachtman, D. P. (2011). 14-3-3 proteins fine-tune plant nutrient metabolism. *FEBS Lett.* 585, 143–147. doi: 10.1016/j.febslet.2010.11.025
- Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., et al. (2019). DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* doi: 10.1093/bioinformatics/bty1054 [Epub ahead of print].
- Singh, S., Vikram, P., Sehgal, D., Burgueño, J., Sharma, A., Singh, S. K., et al. (2018). Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. *Sci. Rep.* 8:12527. doi: 10.1038/s41598-018-30667-4
- Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z
- Tasleem-Tahir, A. (2012). *Proteomic Analysis of Endosperm and Peripheral Layers During Kernel Develop-Ment of Wheat (Triticum aestivum L.) and a Preliminary Approach of Data Integration with Transcriptome*. Agricultural Sciences. Clermont-Ferrand: Université Blaise Pascal.
- 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/pr2011110d
- Uthayakumaran, S., Gras, P. W., Stoddard, F. L., and Beke, F. (1999). Effect of varying protein content and glutenin-to-gliadin ratio on the functional properties of wheat dough. *Cereal Chem.* 76, 389–394. doi: 10.1094/cchem.1999.76.3.389
- Valot, B., Langella, O., Nano, E., and Zivy, M. (2011). MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* 11, 3572–3577. doi: 10.1002/pmic.201100120
- 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, 1594–1611. doi: 10.1002/pmic.200401034
- Victorio, V. C. M., Souza, G. H. M. F., Santos, M. C. B., Vega, A. R., Cameron, L. C., and Ferreira, M. S. L. (2018). Differential expression of albumins and globulins of wheat flours of different technological qualities revealed by nanoUPLC-UDMSE. *Food Chem.* 239, 1027–1036. doi: 10.1016/j.foodchem.2017.07.049
- Vincent, J., Martre, P., Gouriou, B., Ravel, C., Dai, Z., Petit, J.-M., et al. (2015). RulNet: a web-oriented platform for regulatory network inference, application to wheat –omics data. *PLoS One* 10:e0127127. doi: 10.1371/journal.pone.0127127
- Vizcano, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456. doi: 10.1093/nar/gkv1145
- 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. doi: 10.1094/CCHEM.1998.75.5.644
- 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. doi: 10.1002/(sici)1097-0010(199801)76:1<49::aid-jsfa950>3.0.co;2-2
- Xie, Q., Mayes, S., and Sparkes, D. L. (2015). Cernel size, grain filling, and morphology determine individual grain weight in wheat. *J. Exp. Bot.* 66, 6715–6730. doi: 10.1093/jxb/erv378
- Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106. doi: 10.1038/ng.2876
- Yang, M., Gao, X., Dong, J., Gandhi, N., Cai, H., von Wettstein, D. H., et al. (2017). Pattern of protein expression in developing wheat grains identified through proteomic analysis. *Front. Plant Sci.* 8:962. doi: 10.3389/fpls.2017.00962
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Yu, X., Chen, X., Wang, L., Yang, Y., Zhu, X., Shao, S., et al. (2017). Novel insights into the effect of nitrogen on storage protein biosynthesis and protein body development in wheat caryopsis. *J. Exp. Bot.* 68, 2259–2274. doi: 10.1093/jxb/erx108
- Zhao, F. J., Hawkesford, M. J., and McGrath, S. P. (1999). Sulphur assimilation and effects on yield and quality of wheat. *J. Cereal Sci.* 30, 1–17. doi: 10.1006/jcrs.1998.0241
- Zheng, T., Peng, F., Cao, Y. L., Han, Y. N., Ma, H. L., Wang, Y., et al. (2018). Mechanisms of wheat (*Triticum aestivum*). grain storage proteins in response to nitrogen application and its impacts on processing quality. *Sci. Rep.* 8:11928. doi: 10.1038/s41598-018-30451-4

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.

Copyright © 2019 Bancel, Bonnot, Davanture, Alvarez, Zivy, Martre, Déjean and Ravel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.