Cell wall proteome investigation of bread wheat (Triticum aestivum) developing grain in endosperm and outer layers

Mehdi Cherkaoui, Audrey Geairon, Virginie Lollier, Hélène San Clemente, Colette Larré, Hélène Rogniaux, Elisabeth Jamet, Fabienne Guillon, Mathilde Francin-Allami

Dataset Brief

Cell wall proteome investigation of bread wheat (*Triticum aestivum*) developing grain in endosperm and outer layers

Mehdi Cherkaoui[1], Audrey Geairon[1], Virginie Lollier[1], Hélène San Clemente[2], Colette Larré[1], Hélène Rogniaux[1], Elisabeth Jamet[2], Fabienne Guillon[1], Mathilde Francin-Allami[1]

[1] INRA, Biopolymères Interactions Assemblages, Nantes, France

[2] Laboratoire de Recherche en Sciences Végétales, CNRS, UPS, Université de Toulouse, Auzeville, Castanet Tolosan, France

Correspondence : Dr M. Francin-allami, INRA, Biopolymères Interactions Assemblages, Rue de la Géraudière BP 71627, 44316 Nantes Cedex 3, France

E-mail : mathilde.francin-allami@inra.fr

Abbreviations : CE, carbohydrate esterase; **CWP**, cell wall protein; **GH**, glycosyl hydrolase; **PAC**, protein acting on carbohydrates

Keywords : cell wall proteins, grain, polysaccharides, remodeling, wheat

Abstract :

The remodeling of cell wall polysaccharides is controlled by cell wall proteins (CWPs) during the development of wheat grain. This work describes for the first time the cell wall proteomes of the endosperm and outer layers of the wheat developing grain, which have distinct physiological functions and technological uses. Altogether, 636 non-redundant predicted CWPs were identified with 337 proteins in the endosperm and 594 proteins in the outer layers, among which 295 proteins were present in both tissues, suggesting both common and tissue specific remodeling activities. These proteins were distributed into eight functional classes. Approximatively a quarter of them were predicted to act on cell wall polysaccharides, with many glycosylhydrolases and also expansin, lyases and carbohydrate esterases. Therefore, our results provide crucial data to go further in the understanding of relationship between tissue-specific morphogenesis and cell wall remodeling in cereals. Data are available via ProteomeXchange with identifier PXD010367.

Cell walls are extracellular matrices surrounding the cell nearby the plasma membrane and implicated in many physiological roles. They constitute an external physical barrier that protect cells from environmental stresses, confer rigidity to the plant, and allow cell-to-cell adhesion [1]. They are continually remodeled to respond to the physiological needs of the plant during its development. From a technological point of view, cell walls have a major impact on the use of plant biomass for food and non-food applications [2]. Cell walls are mainly composed of polysaccharides (cellulose, hemicelluloses and pectins). In supporting and vascular tissues, they can contain significant amounts of lignin. Proteins are also found and represent about 5-10% of the cell wall dry mass. These cell wall proteins (CWPs) can be structural or enzymatic proteins and they play crucial roles in the assembly of polymer and their remodeling during seed germination, its maturation and development [3,4]. In wheat, a major cereal crop, cell walls only account for about 3–8% of grain dry mass. Nevertheless, the cell wall components have major effects on the end-use of wheat grain (including milling, baking, animal feed) and constitute a source of dietary fiber in cereal foods [5]. In this context, considerable efforts have been made to improve our understanding of the structure and properties of grain cell wall components.

Wheat grain is constituted of several parts with specific physiological functions: the embryo, the endosperm and the outer layers. Endosperm accumulates storage products for sustaining the early seedling growth while the outer layers comprising the pericarp, the seed coat and the nucellus epidermis, fulfill a nutritive function at the early stage of grain development and then a protective role. Cell walls in the outer layers are mainly formed of cellulose and complex xylans but they also contain a small amount of lignin while in the endosperm they are mainly composed of arabinoxylans and mixed β-glucan [6–9]. To further understand cell wall dynamics during grain development it is necessary to study how polysaccharides are built and remodeled *in muro*. As mentioned above, these events are regulated by CWPs [4]. In order to identify them, sub-cellular proteomics targeting the cell wall must be implemented.

Cell wall proteomes of various plant species have been investigated. To date, *Arabidopsis thaliana* has the best documented with several sub-proteomes published [10–14]. Cell wall proteomes of some

monocot species like *Oryza sativa, Brachypodium distachyon* or *Saccharum officinarum* have also been investigated but few studies have focused on their grains [15–17]. Recently, the cell wall proteome of the *B. distachyon* grain, a model plant for *Poaceae*, has been published [18,19]. The data were obtained at different stages of grain development, giving a dynamic view of cell wall organization. However, cell walls in *B. distachyon* grain greatly differ from those of wheat grain. They are much thicker, associated with high mixed β-glucan content which could also serve as a storage carbohydrate to supplement the low content of starch in the grain of *B. distachyon* (in addition to its structural role). Mixed β-glucans are enriched in (1→3) linkages and arabinoxylans are highly feruloylated when compared with wheat, barley, or oat [20–22]. For these reasons, the events occurring in the cell walls of *B. distachyon* and wheat grains cannot be directly compared.

To gain more information about the mechanisms of cell wall assembly and remodeling in cultivated *Poaceae*, we studied the cell wall proteome of the wheat developing grain. The endosperm and the outer layers of grains were separately investigated as we presume they will have different protein composition to fulfill their different functions [7].

Triticum aestivum cv. Recital was grown in a greenhouse under natural day-length conditions. The development of plants was calculated based on growing degree-day (GDD). Plants were harvested at 250 GDD after flowering. This stage corresponds to the end of cellularization and to the beginning of grain filling [7]. At this stage, cell walls are completely synthetized, cell division in the endosperm is ended and storage compounds start to accumulate [8]. At the same time, in the outer layers, mesocarp cells undergo lysis, whereas cells of the seed coat and nucellar epidermis also degenerate but keep their walls. Grains were manually dissected to separate outer layers and endosperm. Protein extracts were obtained as described previously [18]. Briefly, the protocol includes two major steps: first, the sub-cellular fractionation of cell walls by centrifugation in increasing sucrose concentration; second, the extraction of proteins by successive washes in acetate buffer 5mM pH 4.6 supplemented with 0.2 M $CaCl_2$ or 2 M LiCl. Five biological replicates were prepared for both endosperm and outer layers in order to maximize robustness of the experiment and the number of identified proteins. Starting from 500 mg of fresh material, we obtained an average of 40 mg and 60 mg of purified cell walls from

endosperm and outer layers respectively. Protein extraction allowed us collecting an average of 0.5 mg (endosperm) and 0.8 mg (outer layers) of extracts enriched in CWPs. The quality of the protein extracts was confirmed by 1D-electrophoresis (1-DE). Twenty µg of each sample were analyzed in precast-gels (Mini-PROTEAN TGX Stain-Free 12% Precast Gels, BIORAD) and stained with Coomassie blue (Figure 1). For both endosperm and outer layers' protein extracts, the profiles look very similar between the five biological replicates, thus showing a good reproducibility of the experiment. After this control, 50 µg of each sample were briefly separated by 1D-E in precast gels. Then, each lane was cut in five fragments prior to an "in-gel" tryptic digestion, according to Suliman *et al.* [23].

The trypsin-digested samples (6 µL) were injected into the LC-MS/MS system. The mass analyses were performed using an LTQ-Orbitrap VELOS mass spectrometer (*i.e.* a Linear ion Trap Quadrupole mass filter associated with an Orbitrap™ analyzer, Thermo-Fisher Scientific) coupled to a nanoscale LC system (Ultimate U3000 RSLC system, Thermo-Fisher Scientific). Chromatographic separation was performed on a reversed-phase capillary column (Acclaim® PepMap™ C18 2 µm 100 A, 75 µm i.d. x 50 cm long, Thermo-Fisher Scientific/Dionex) at 60°C using a linear gradient performed between the following mobile phases: (A) 97.9% water, 2% ACN, 0.1% TFA; (B) 90% ACN, 0.08% formic acid. The other parameters were as described in [23]. Analyses were performed using a typical survey method, in which full MS scans were acquired at 60 000 resolution (FWMH) using the Orbitrap analyzer (m/z 300–2000) while the CID spectra for eight most intense ions were recorded in the LTQ. The LC-MS/MS raw data were processed into mzXML files and were further searched against databases using the X!Tandem pipeline software (http://pappso.inra.fr/bioinfo/xtandempipeline/) [24].

Protein identification was achieved by comparing the MS data to the UniProt databank restricted to *T. aestivum* (release May 2017, 136,892 accessions, http://www.uniprot.org/). Parameters, peptide identification results, protein inference and false discovery rate are detailed in Supporting Information Tables S1, S2, S3 and S4. Briefly, protein identification was validated when at least two specific peptides were found in a given sample. To deal with the strong redundancy of identifications due to

the three copies of bread wheat genome [25], we decided to cluster the identified proteins with 90% amino acid sequence identity (Supporting Information Table S5). The protein with the longest amino acid sequence was chosen to represent the cluster. In addition, we only validated the protein clusters which have members present at least in two of the five biological replicates. In this way, we produced two sets of non-redundant CWPs representing the cell wall enriched proteome of both parts of the grain, the endosperm and the outer layers. The overall workflow of the analysis is detailed in Supporting Information Figure S1.

The *ProtAnnDB* pipeline was used to annotate these proteomes (http://www.polebio.lrsv.ups-tlse.fr/ProtAnnDB/index.php) [26]. It gives access to multiple bioinformatic programs to predict the sub-cellular localization of proteins, and functional domains. In addition, a bioinformatic analysis was done with the Phobius software (http://www.phobius.sbc.su.se/). The assignation of a protein as a CWP was performed by successive steps. Firstly, only the proteins displaying predictions of a signal peptide by at least two bioinformatic programs were retained and classified into a list of "secreted proteins". Secondly, all proteins displaying an ER retention signal were rejected and classified into an "intracellular proteins" group. The last criterion was based on the analysis of functional domain predictions of proteins and allowed us to define a CWP group from the "secreted proteins". Biological interpretation of data was achieved by analysis of predicted functional domains. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD010367 and 10.6019/PXD010367".

A clustering analysis was achieved from the binary distance matrix computation using the R framework (v. 3.4.0, www.rproject.org/) in order to evaluate similarity and reproducibility between samples. Consistency of clusters was assessed by bootstrap values with the help of the pvclust library [27] (Figure 2).

As expected, the samples are conveniently clustered according to the grain parts. There was no difference between bootstraps realized from all identified proteins and from CWPs (Supporting Information Fig. S2) thus allowing to validate the quality of sample preparation.

Altogether, a set of 2905 non-redundant proteins was identified: 2121 were found in the endosperm and 2315 in the outer layers. The bioinformatics analysis described above allowed defining 636 non-redundant proteins predicted as CWPs. Among them, 337 and 594 CWPs were found in the endosperm and the outer layers, respectively (Supporting Information Table S6). This corresponded to a proportion of CWPs of 15.8% for endosperm and 25.6% for outer layers. These yields are slightly lower than those of previous studies using similar protocols for cell wall enrichment and protein extraction [18,19]. We assume that the amount of intracellular proteins (Supporting Information Table S7) was particularly high in the endosperm samples probably due to their low proportion of cell walls at this developmental stage. While our cell wall enrichment seems to be less efficient than in previous studies, the total number of identified CWPs is slightly higher than that obtained on the whole grain of *B. distachyon* at a similar developmental stage [18].

The comparison between the CWPs identified in endosperm and outer layers of the grain shows that almost half of the proteins (295) are present in both of them, suggesting common cell wall metabolism activities. Forty-two proteins were only identified in the endosperm, whereas 299 were only found in the outer layers. This may be explained by the fact that outer layers are composed of different tissues compared to the endosperm, which implies a greater diversity of polysaccharides, and so the requirement of more CWPs.

All the identified CWPs were classified into eight functional classes usually used to sort CWPs according to their functional domain prediction (Figure 3). Distribution of CWPs into these classes was very similar between grain endosperm and outer layers. The two most represented classes are proteins acting on cell wall polysaccharides (PAC), and proteases. As commonly observed in other cell wall proteomes, PAC represent about 23% of all identified CWPs [13]. The PAC class mainly contains glycosyl hydrolases (GHs), expansins, lyases or carbohydrates esterases (CEs), which display prominent roles in cell wall polysaccharide assembly and remodeling [4]. It was noticed that protease abundance is slightly higher in the endosperm than in the outer layers (21% and 18%, respectively). The protease class contains proteins of the subtilisin family like Ser carboxypeptidases, Asp proteases or Ser proteases that fulfill highly specific functions in plant development and signaling cascades [28–

30]. Similar to proteases, proteins with interacting domains were more represented in the endosperm (20%) than in the outer layers of wheat grain (14%). This class is mostly represented by putative plant invertase or pectin methylesterase inhibitors which are critical for cell wall dynamics [31]. It can be noticed that these latter are less represented among the outer layers specific CWPs (Supplementary Information Table S1). Altogether, these three functional classes represent more than a half of the total number of CWPs identified in the endosperm and the outer layers of wheat grains. Proteins related to lipid metabolism represent also a large proportion of the identified proteins, especially in the endosperm, where they account for 11% in the endosperm *vs* 14% in the outer layers. This functional class mainly contains lipid transfer proteins (LTPs) which were described to have, among others, wall-loosening activity [32]. The class of oxido-reductases (7% in endosperm and 10% in outer layers) mainly contains class III peroxidases. These proteins are highly conserved and diversified in land plants and are key players of cell wall modification [33]. They are not represented among the CWPs specific to endosperm. We identified in the same proportions miscellaneous CWPs and proteins with unknown functions (7% and 9% respectively in endosperm and 8% for both in outer layers) what is classically found in other cell wall proteomes [13,34,35]. Finally, the less represented CWPs in our dataset are those involved in signaling which accounts for 2% in the endosperm and 4% in the outer layers, No structural protein was identified in this analysis: this could be explained by their low abundance in the cell wall of monocot plants and/or their difficulties to be extracted [13,36].

To our knowledge, this is the second description of the cell wall proteome of a monocot. The first one was that of the *B. distachyon* developing grain [18,19]. However, this work describes the first cell wall proteome discriminating two parts of the grain, the endosperm and the outer layers. It gives information on cell wall metabolism of these two parts of the grain which have distinct physiological functions and technological uses [7]. Numerous CWPs, specific or not to endosperm or outer layers, were identified. Among them almost a quarter are enzymes potentially catalyzing reactions involving cell wall polysaccharides (such as GHs, CEs, lyases and peroxidases) but so far their exact functions are still poorly known. They are potential candidates to be characterized in order to better understand cell wall polysaccharide assembly and remodeling during wheat grain development.

Acknowledgements

This work was supported by INRA funding. EJ and HSC are thankful to CNRS and the Paul Sabatier-Toulouse 3 University for supporting their research.

The authors have declared no conflict of interest.

[1] H. Le Gall, F. Phillipe, J. M. Domon, F. Gillet, J. Pelloux, C. Rayon, Plants 2015, 4, 112.

[2] G. Fincher, B. Stone, Adv. Cereal Technol. 1986, 8.

[3] G. Fincher, Plant Physiol. 2009, 149, 27.

[4] L. Franková, S. C. Fry, J. Exp. Bot. 2013, 64, 3519.

[5] G. McDougall, I. Morrison, D. Stewart, R. Hillman, J. Sci. Food Agric. 1996, 70, 133.

[6] C. Barron, A. Surget, X. Rouau, J. Cereal Sci. 2007, 45, 88.

[7] L. Saulnier, F. Guillon, A. L. Chateigner-Boutin, J. Cereal Sci. 2012, 56, 91.

[8] S. Philippe, L. Saulnier, F. Guillon, Planta 2006, 224, 449.

[9] L. Saulnier, P. Robert, M. Grintchenko, F. Jamme, B. Bouchet, F. Guillon, J. Cereal Sci. 2009, 50, 312.

[10] C. Albenne, H. Canut, E. Jamet, *Front. Plant Sci.* 2013, 4, 1.

[11] U. Basu, J. Francis, R. Whittal, J. Stephens, Y. Wang, O. Zaiane, R. Goebel, D. Muench, A. Good, G. Taylor, Plant Soil 2006, 286, 357.

[12] V. Hervé, H. Duruflé, H. San Clemente, C. Albenne, T. Balliau, M. Zivy, C. Dunand, E. Jamet, Proteomics 2016, 16, 3183.

[13] H. Duruflé, H. San Clemente, T. Balliau, M. Zivy, C. Dunand, E. Jamet, Proteomics 2017, 17, 1.

[14] A. Y. Tsai, T. Kunieda, J. Rogalski, L. J. Foster, B. E. Ellis, G. W. Haughn, Plant Physiol. 2017, 173, 1059.

[15] W. K. Cho, X. Y. Chen, H. Chu, Y. Rim, S. Kim, S. T. Kim, S. W. Kim, Z. Y. Park, J. Y. Kim, Physiol. Plant. 2009, 135, 331.

[16] T. Douché, H. San Clemente, V. Burlat, D. Roujol, B. Valot, M. Zivy, R. Pont-Lezica, E. Jamet, Proteomics 2013, 13, 2438.

[17] J. Fonseca, M. Calderan-Rodrigues, F. de Moraes, T. Cataldi, E. Jamet, C. Labate, Proteomics 2018, 18, 10.

[18] M. Francin-Allami, K. Merah, C. Albenne, H. Rogniaux, M. Pavlovic, V. Lollier, R. Sibout, F. Guillon, E. Jamet, C. Larré, Proteomics 2015, 15, 2296.

[19] M. Francin-Allami, V. Lollier, M. Pavlovic, H. San Clemente, H. Rogniaux, E. Jamet, F.

Guillon, C. Larré, Proteomes 2016, 4, 21.

[20] M. Opanowicz, P. Hands, D. Betts, M. L. Parker, G. Toole, E. N. C. Mills, J. H Doonan, S. Drea, J. Exp. Bot. 2011, 62, 735.

[21] F. Guillon, B. Guillon, F. Jamme, P. Robert, B. Quéméner, C. Barron, C. Larré, P. Dumas, L. Saulnier, J. Exp. Bot. 2011, 62, 1001.

[22] F. Guillon, C. Larré, F. Petipas, A. Berger, J. Moussawi, H. Rogniaux, A. Santoni, L. Saulnier, F. Jamme, M. Miquel, L. Lepiniec, B. Dubreucq, J. Exp. Bot. 2012, 63, 739.

[23] M. Suliman, A. L. Chateigner-Boutin, M. Francin-Allami, A. Partier, B. Bouchet, J. Salse, C. Pont, J. Marion, H. Rogniaux, D. Tessier, F. Guillon, C. Larré, J. Proteomics 2012, 78, 508.

[24] O. Langella, B. Valot, T. Balliau, M. Blein-Nicolas, L. Bonhomme, M. Zivy, J. Proteome Res. 2017, 16, 494.

[25] Y. Matsuoka, Plant Cell Physiol. 2011, 52, 750.

[26] H. San Clemente, R. Pont-Lezica, E. Jamet, Bioinform. Biol. Insights 2009, 15.

[27] R. Suzuki, H. Shimodaira, Bioinformatics 2006, 22, 1540.

[28] R. Siezen, J. Leunissen, Protein Sci. 2008, 6, 501.

[29] A. Figueiredo, F. Monteiro, M. Sebastiana, Front. Plant Sci. 2014, 5, 739.

[30] R. van der Hoorn, Annu. Rev. Plant Biol. 2008, 59, 191.

[31] A. Peaucelle, S. Braybrook, H. Höfte, Front. Plant Sci. 2012, 3, 121.

[32] J. Nieuwland, R. Feron, B. Huisman, A. Fasolino, C. Hilbers, J. Derksen, C. Mariani, Plant Cell 2005, 17, 2009.

[33] E. Francoz, P. Ranocha, H. Nguyen-Kim, E. Jamet, V. Burlat, C. Dunand, Phytochemistry 2015, 112, 15.

[34] A. Day, S. Fénart, G. Neutelings, S. Hawkins, C. Rolando, C. Tokarski, Proteomics 2013, 13, 812.

[35] M. J. Calderan-Rodrigues, E. Jamet, T. Douché, M. B. Bonassi, T. Cataldi, J. Fonseca, H. San Clemente, R. Pont-Lezica, C. Labate, BMC Plant Biol. 2016, 16, 1.

[36] J. Vogel, Curr. Opin. Plant Biol. 2008, 11, 301.

Figures

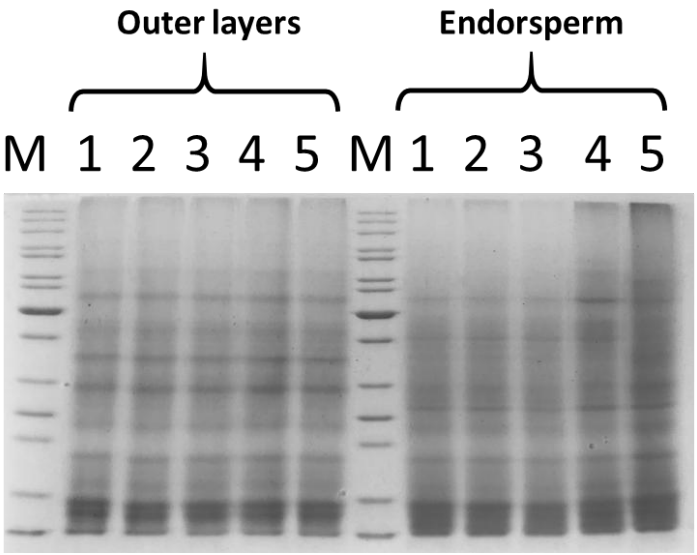

Figure 1: 1-DE profiles of wheat grain proteins extracted from enriched cell walls fractions. Twenty µg of each sample (#1-#5: biological replicates 1-5, respectively) have been separated by 1-DE and stained with Coomassie blue. Molecular mass markers (M) are in kDa.

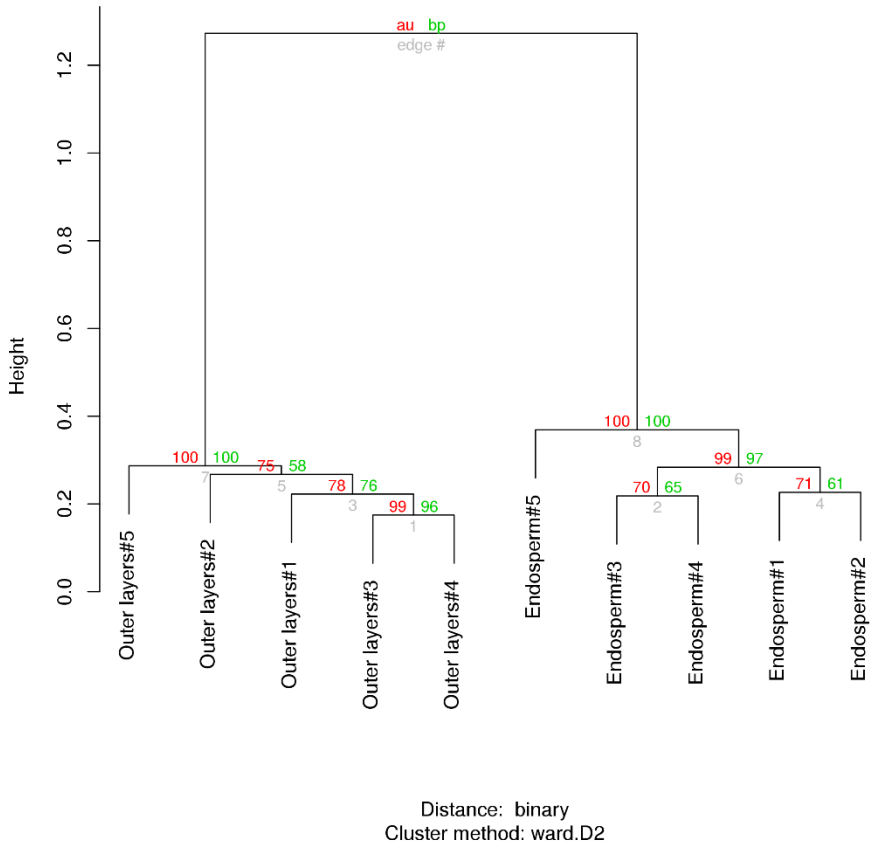

Figure 2 : Hierarchical clustering of the 10 samples from the identified CWPs. Twice five biological replicates (#1-5) from outer layers or endosperm of the grain are taken into account. Bootstrap probabilities (bp) are printed in green, approximately unbiased p-values (au) in red and edge numbers in gray.

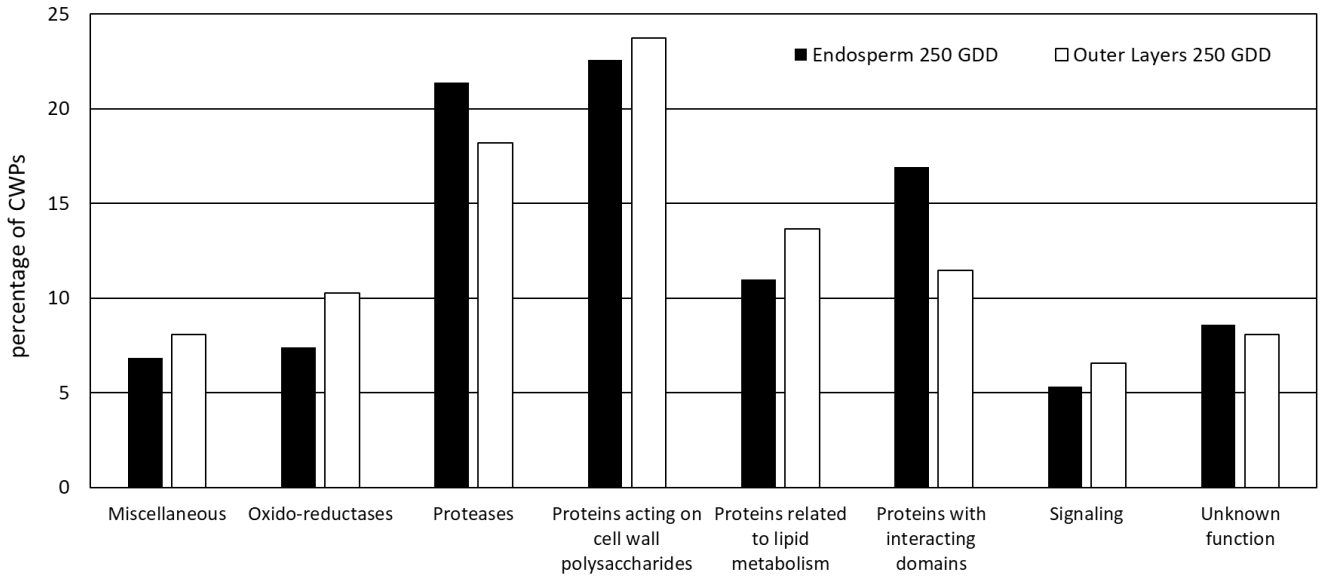

Figure 3: Distribution of the identified CWPs in functional classes in the cell wall proteomes of the endosperm and the outer layers of the wheat grain. M= Miscellaneous ; PI= Proteins with interacting domains ; OR= Oxido-Reductases ; P= Proteases ; U= Unknown function ; PAC= Proteins acting on cell wall polysaccharides ; S= Signaling ; PL= Proteins related to lipid metabolism