



ELSEVIER

Contents lists available at ScienceDirect

## Data in brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

## Dataset of the metabolic and CM-like protein fractions in old and modern wheat Italian genotypes



Antonella Di Francesco <sup>a</sup>, Rosaria Saletti <sup>a, \*</sup>,  
 Vincenzo Cunsolo <sup>a</sup>, Birte Svensson <sup>b</sup>, Vera Muccilli <sup>a</sup>,  
 Pasquale De Vita <sup>c</sup>, Salvatore Foti <sup>a</sup>

<sup>a</sup> Laboratory of Organic Mass Spectrometry, Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125, Catania, Italy

<sup>b</sup> Department of Biotechnology and Bioengineering, Technical University of Denmark, Søtofts Plads, Building 224, Kgs. Lyngby, DK-2800, Denmark

<sup>c</sup> CREA Research Centre for Cereal and Industrial Crops (CREA-CI), S.S. 673 km 25.200, 71122, Foggia, Italy

## ARTICLE INFO

## Article history:

Received 3 July 2019

Received in revised form 19 October 2019

Accepted 21 October 2019

Available online 1 November 2019

## Keywords:

Old and modern wheat genotypes

High resolution mass spectrometry

Proteome analysis

Proteins

Allergens

Food and nutrition

## ABSTRACT

The present work reports the first comprehensive proteomic profiling and qualitative comparison of metabolic and Chloroform-Methanol (CM)-like protein fractions extracted from mature kernels of two old Sicilian durum wheat landraces, *Russello* and *Timilia Reste Bianche* (*Timilia RB*), and *Simeto*, an improved durum wheat variety widespread in Italy and other Mediterranean countries and chosen as representative of the most widely commercial cultivars. The data are discussed in the related research article “Qualitative proteomic comparison of metabolic and CM-like protein fractions in old and modern wheat Italian genotypes by a shotgun approach” [1]. The results of this work could be used for *in vitro* investigations to understand the relationship between protein profiles of old and modern wheat genotypes and their potential benefits for human consumption.

© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Corresponding author.

E-mail address: [rsaletti@unict.it](mailto:rsaletti@unict.it) (R. Saletti).

Specifications Table

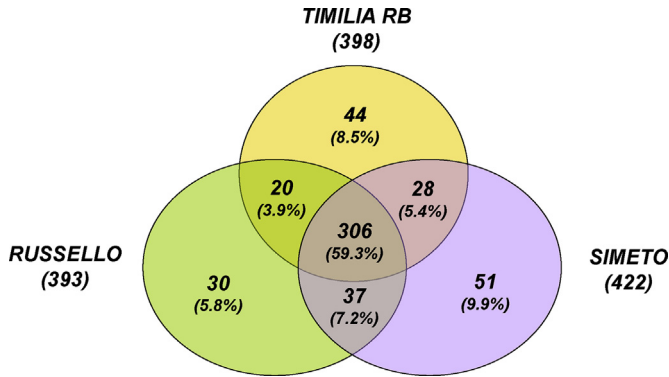
Subject	Food Science
Specific subject area	Protein profile of old and modern durum wheat genotypes
Type of data	Graph Figure
How data were acquired	Raw data UHPLC/High-resolution mass spectrometry Instrument: Thermo Fisher Scientific Orbitrap Fusion Tribrid® (Q-OT-qIT) mass spectrometer Software: <i>Xcalibur</i> v. 3.0.63 software (Thermo Fisher Scientific), PEAKS de novo sequencing software (v. 8.5, Bioinformatics Solutions Inc., Waterloo, ON Canada).
Data format	Raw Analyzed
Parameters for data collection	Three biological replicates of <i>Russello</i> , <i>Timilia RB</i> and <i>Simeto</i> were provided from CREA-CI. The genetic materials were sowed at Foggia, during the 2010-11 growing season, following a randomized block design with three replicates. Grain samples were harvested and the flours were stored at 4 °C.
Description of data collection	Flours (200 mg) were suspended in 2 mL cold (4 °C) extraction solution (50 mM Tris-HCl, 100 mM KCl, 5 mM EDTA, Protease Inhibitor Cocktail, pH 7.8) in order to obtain the metabolic and CM-like proteins [2]. Each extract were reduced by adding 40 µg of DTT, alkylated with 96 µg of IAA and digested by porcine trypsin at an enzyme-substrate ratio of 1:50 (overnight, 37 °C) [3]. Each peptide mixture was subjected to triplicate RP-nHPLC/nESI-MS/MS analyses and database search.
Data source location	Department of Chemical Sciences, Laboratory of Organic Mass Spectrometry, University of Catania, Catania, Italy.
Data accessibility	Repository name: PRIDE via ProteomeXchange Data identification number: PXD014449 <a href="http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD014449">http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD014449</a>
Related research article	Antonella Di Francesco, Rosaria Saletti, Vincenzo Cunsolo, Birte Svensson, Vera Muccilli, Pasquale De Vita and Salvatore Foti, <b>Qualitative proteomic comparison of metabolic and CM-like protein fractions in old and modern wheat Italian genotypes by a shotgun approach</b> , J. Proteom [1]. <a href="https://doi.org/10.1016/j.jprot.2019.103530">https://doi.org/10.1016/j.jprot.2019.103530</a>

### Value of the Data

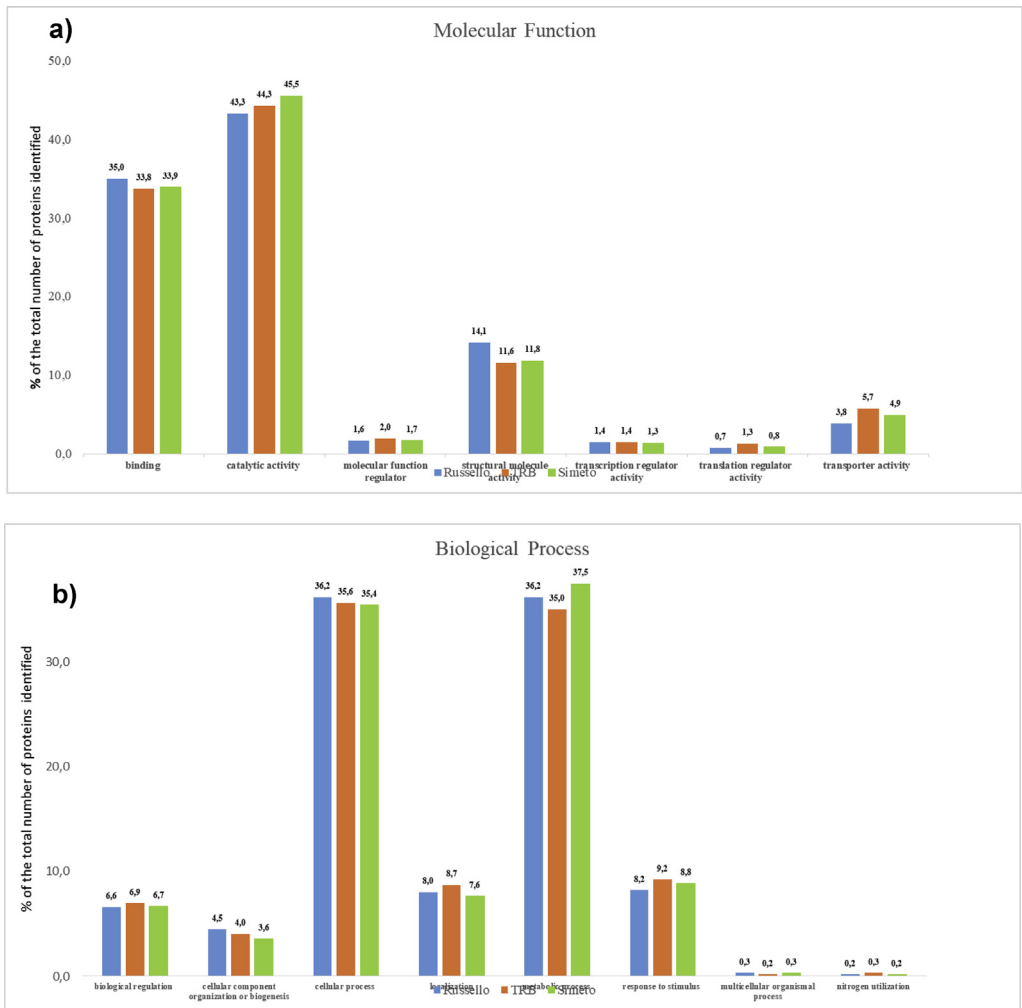
- These data are useful for elucidating the relationship between protein profiles of old and modern wheat genotypes and their potential benefits for human consumption.
- These data can be useful for further studies on the relationship between metabolic protein profile of wheat and allergic properties such as celiac disease (CD), wheat allergy and non-celiac wheat sensitivity (NCWS).
- *In vitro* investigations of the allergic properties of the metabolic extracts of old and modern genotypes may benefit of the knowledge of the protein profiles.
- The present work reports the first comprehensive proteomic profiling and qualitative comparison of metabolic and CM-like protein fractions of old and modern wheat genotypes.

## 1. Data

A pairwise comparison of the metabolic fractions of the three varieties investigated was carried out (Fig. 1). In detail, comparison of *Russello* and *Simeto* shows that these two genotypes share 343 unique gene products, whereas 50 and 79 were exclusively identified in the metabolic fractions of *Russello* and *Simeto*, respectively (Fig. 1a). Gene Ontology (GO) analysis was limited to the unique gene products identified exclusively in *Russello* and *Simeto* and allowed to ascertain the molecular function and biological process in which they are involved (Fig. 2). As shown in Fig. 2a the proteins found exclusively in each of these two genotypes are mainly involved in “catalytic activity” and “binding”. Only minor differences were observed in the “structural molecule activity”, found only in the modern cultivar, the “transporter activity” more represented in *Simeto* and the “molecular function regulator” more represented in *Russello*. Also, the comparison of biological process distribution (Fig. 2b) does not show significant differences for almost all the processes reported, with two exceptions, “response to



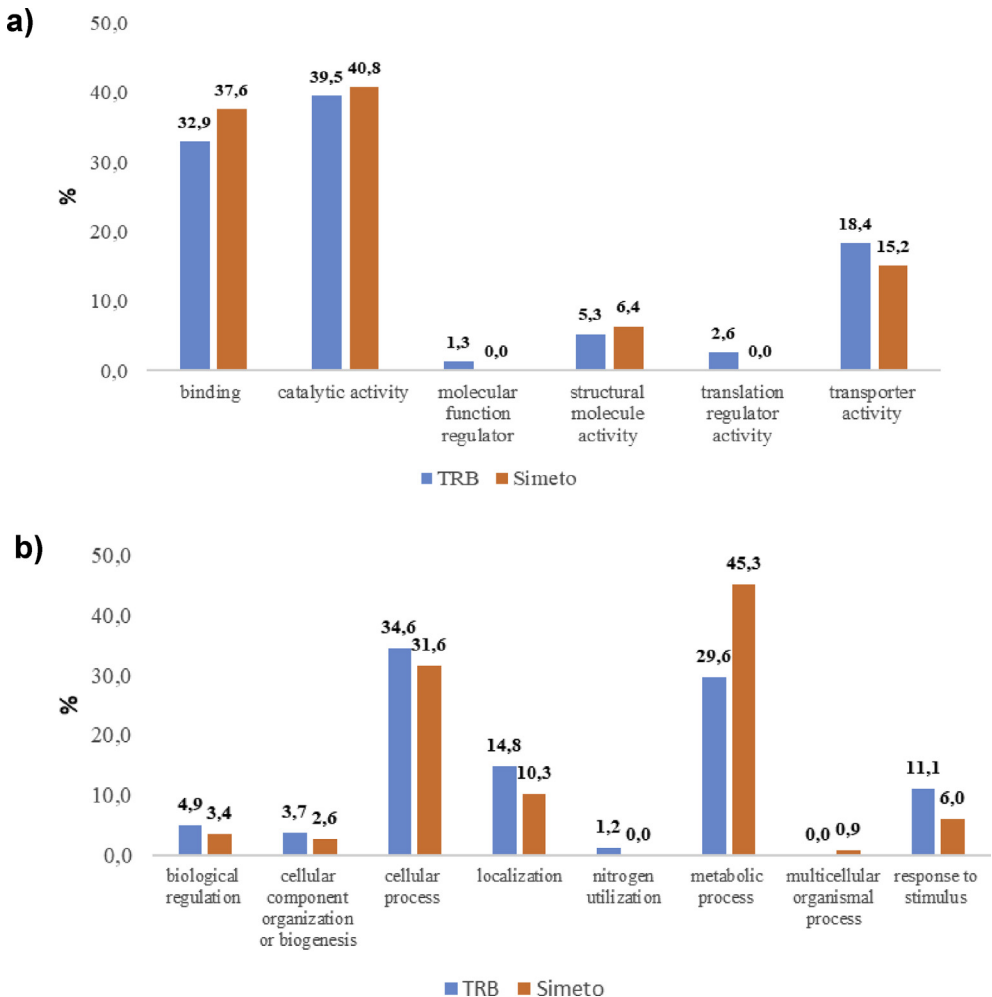
**Fig. 1.** Pairwise comparison of the proteins identified in the metabolic fractions of the three genotypes investigated.



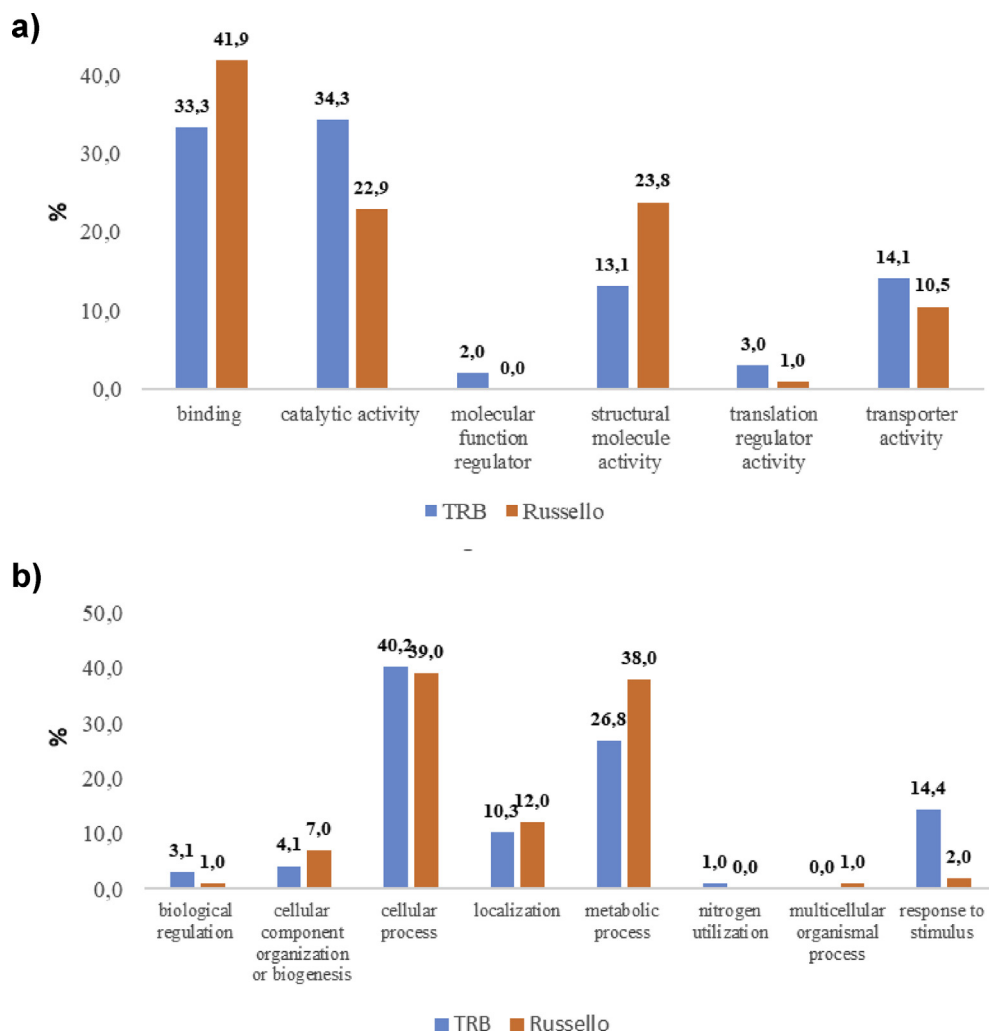
**Fig. 2.** Histograms of (a) molecular functions, and (b) biological processes pertaining to the exclusive unique gene products detected in *Russello*, with respect to *Simeto* and viceversa.

stimulus” that includes more exclusive proteins in *Simeto* compared to *Russello*, and “cellular component organization or biogenesis” and “developmental process” more represented in the old genotype.

The lists of proteins identified in the metabolic fraction of *Timilia RB* and *Simeto* share 334 proteins (Fig. 1b). In contrast, 64 proteins were unique for *Timilia RB* and 88 for *Simeto*. Molecular function and biological process distributions of the unique proteins in each of these two genotypes are reported in Fig. 3. The main molecular functions are “binding”, “catalytic activity” and “transporter activity” (Fig. 3a). In addition, there are some particular differences, genotype *Simeto* does not present exclusive proteins playing “molecular function regulator” and “translation regulator activity”. The distribution of the biological process (Fig. 3b) shows the general trend already observed in the comparison of *Russello* and *Simeto* with some differences. The “developmental process” in this comparison is not represented and the “response to stimulus” does not show differences between *Timilia RB* and *Simeto*. Moreover, the “nitrogen utilization” process is unique of the old genotype and, on the contrary, “multicellular organismal process” is unique of *Simeto*.



**Fig. 3.** Histograms of (a) molecular functions, and (b) biological processes pertaining to the exclusive unique gene products detected in *Timilia RB*, with respect to *Simeto* and *viceversa*.



**Fig. 4.** Histograms of (a) molecular functions, and (b) biological processes pertaining to the exclusive unique gene products detected in *Timilia RB*, with respect to *Russello* and *viceversa*.

Finally, the comparison of the proteins identified in the metabolic fraction of the two old genotypes *Timilia RB* and *Russello* (Fig. 1c) shows they share 326 unique gene products, while 72 proteins are unique for *Timilia RB* and 67 for *Russello*. GO classification of the molecular function (Fig. 4a) reveals minor differences, the “translational regulator activity” is more represented in *Timilia RB* and “molecular function regulator” is not found in *Simeto*. The biological process (Fig. 4b) have the same distribution with some differences. The “multicellular organismal process” is represented only in *Russello*, on the contrary, *Russello* does not present exclusive proteins playing “nitrogen utilization”. Moreover, the number of exclusively proteins involved in the process “response to stimulus” and “biological regulation” is higher in *Russello*.

## 2. Experimental design, materials, and methods

### 2.1. Chemicals

All chemicals were of the highest purity commercially available and were used without further purification. KCl, K<sub>2</sub>HPO<sub>4</sub>, acetone, methanol, acetic acid and Tris-HCl were purchased from Carlo Erba (Milan, Italy). Formic Acid (FA), Protease Inhibitor Cocktail, EDTA, ammonium bicarbonate, ammonium acetate, dithiothreitol (DTT), iodoacetamide (IAA) were obtained from Aldrich (St. Louis, Missouri, USA). Modified porcine trypsin was purchased from Promega (Madison, WI, USA). Water and acetonitrile (ACN) (OPTIMA® LC/MS grade) for LC/MS analyses were purchased from Fisher Scientific (Milan, Italy). LDS sample buffer, Mark12™ Unstained Standards and SimplyBlue™ Safe Stain were obtained from Invitrogen™ (Life Technologies™, Paisley, UK).

### 2.2. Samples collection and treatment

Three biological replicates of *Russello*, *Timilia RB* and *Simeto* were provided from CREA-Cl. The genetic materials were sowed at Foggia, during the 2010–11 growing season, following a randomized block design with three replicates. Grain samples were harvested and the flours were stored at 4 °C. Flours (200 mg) were suspended in 2 mL cold (4 °C) extraction solution (50 mM Tris-HCl, 100 mM KCl, 5 mM EDTA, Protease Inhibitor Cocktail, pH 7.8) in order to obtain the metabolic and Chloroform-Methanol (CM)-like proteins [2]. The solution was incubated on ice (5 min) with intermittent mixing and centrifuged (13523 g, 15 min, 4 °C). The supernatant was collected and added five volumes of 0.1 M ammonium acetate in methanol. Following incubation overnight at –20 °C, the solution was centrifuged (30 g, 15 min, R.T.). The pellet (containing the metabolic proteins) was collected and rinsed in 3 mL 0.1 M ammonium bicarbonate, pH 8.2. Proteins in the supernatant (CM-like proteins) were precipitated by addition of four volumes of cold acetone, kept overnight at –20 °C and subsequently centrifuged (30 g, 15 min, R.T.). Finally, the pellet containing CM-like proteins was rinsed by 1.5 mL 0.1 M ammonium bicarbonate, pH 8.2. The protein concentration for each extract was determined by a fluorimetric assay using the Qubit Protein Assay kit with the Qubit 1.0 Fluorometer (ThermoFisher Scientific, Milan, Italy) [3]. Finally, 40 µL (corresponding to about 60 µg) of each extract were reduced by adding 40 µg of DTT dissolved in the same buffer (3 h, 20 °C), alkylated with 96 µg of IAA (1 h, in the dark at 20 °C) and digested by porcine trypsin (Sequencing Grade Modified Trypsin, Porcine, lyophilized, Promega) at an enzyme-substrate ratio of 1:50 (overnight, 37 °C) [3].

### 2.3. Mass spectrometry analysis

Mass spectrometry data were acquired on a Thermo Fisher Scientific Orbitrap Fusion Tribrid® (Q-OT-qIT) mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Liquid chromatography was carried out using a Thermo Scientific Dionex UltiMate 3000 RSLCnano system (Sunnyvale, CA). One microliter of peptide mixture was loaded onto an Acclaim® Nano Trap C18 Column (100 µm i.d. x 2 cm, 5 µm particle size, 100 Å). After washing the trapping column with solvent A (H<sub>2</sub>O + 0.1% FA) for 3 min at a flow rate of 7 µL/min, the peptides were eluted from the trapping column onto a PepMap® RSLC C18 EASY-Spray column (75 µm i. d. x 50 cm, 2 µm particle size, 100 Å) and separated by elution at a flow rate of 0.25 µL/min at 40 °C by a linear gradient of solvent B (ACN + 0.1% FA) in A, 5% for 3 min, followed by 5%–20% in 32 min, 20%–40% in 30 min, 40%–60% in 20 min and 60%–98% in 15 min, finishing by holding 98% B 5 min, 98%–5% in 1 min. and re-equilibrating at 5% B for 20 min. The eluting peptide cations were converted to gas-phase ions by electrospray ionization using a source voltage of 1.75 kV and introduced into the mass spectrometer through a heated ion transfer tube (275 °C). Survey scans of peptide precursors from 200 to 1600 *m/z* were performed at 120 K resolution (@ 200 *m/z*). Tandem MS was performed by isolation at 1.6 Th with the quadrupole, HCD fragmentation with a normalized collision energy of 35, and rapid scan MS analysis in the linear ion trap (low resolution MS/MS analysis). Only those precursors with charge state 2 ÷ 4 and intensity above the threshold of 5 · 10<sup>3</sup> were sampled for MS<sup>2</sup>. The dynamic exclusion duration was set to 60 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The

instrument was run in top speed mode with 3 s cycles, meaning it would continuously perform MS<sup>2</sup> events until the list of non-excluded precursors diminished to zero or 3 s, whichever is shorter. MS/MS spectral quality was enhanced enabling the parallelizable time option (i.e. by using all parallelizable time during full scan detection for MS/MS precursor injection and detection). Mass spectrometer calibration was performed by using the Pierce® LTQ Velos ESI Positive Ion Calibration Solution (Thermo Fisher Scientific). MS data acquisition was carried out by utilizing the *Xcalibur v. 3.0.63* software (Thermo Fisher Scientific).

#### 2.4. Database search, protein identification and gene ontology analysis

MS data were processed using PEAKS de novo sequencing software (v. 8.5, Bioinformatics Solutions Inc., Waterloo, ON Canada). Data were searched against a dedicated protein database (7612 protein sequences), including only the reviewed entries of *Triticum*, *Oryza*, *Hordeum*, *Avena*, *Secale*, *Maize* and *Brachypodium* species downloaded from the UniProt database (release July 2018). The common Repository of Adventitious Proteins (c-RAP) contaminant database was included in the database search.

Database search was carried out using the following parameters: i) full tryptic peptides with a maximum of 3 missed cleavage sites; ii) cysteine carbamidomethylation as a fixed modification; iii) oxidation of methionine, the transformation of N-terminal glutamine and N-terminal glutamic acid residue to pyroglutamic acid form as variable modifications. The precursor mass tolerance threshold was set to 10 ppm and the maximum fragment mass error was set to 0.6 Da. Peptide spectral matches (PSM) were validated using a Target Decoy PSM Validator node based on q-values at a 0.1% False Discovery Rate (FDR). PEAKS score thresholds for Peptide spectral matches (PSMs) were set in order to achieve for each database search FDR values for PSMs, Peptide sequences and Proteins identified below the 0.1% value. This resulted in a range for PEAKS score thresholds from 39 to 47. A protein was considered identified if a minimum of two peptides were matched. Proteins containing the same peptides and that could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony (groups of parsimony). In these cases, proteins from *Triticum*, when identified, were always chosen as the group's reference protein. When a group of parsimony did not contain a component from *Triticum*, the reference protein was selected from the species closest related to *Triticum* and represented in the group.

The gene code, when available, was assigned to the proteins here identified by using the UniProt Knowledge database (<http://www.uniprot.org/>). It should be noted that, due to the limited annotation of the wheat proteins, to obtain coding gene information an additional step was required in many cases. Actually, when the gene symbol was not available, the corresponding protein sequence was subjected to a sequence similarity search by BLAST (Basic Local Alignment Search Tool; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>). By this strategy, many of the identified proteins whose gene code was not available were classified by finding homologous proteins from the species closest related to *Triticum* present in the databases and sharing at least more than 70% sequence similarity. Subsequently, to carry out the Gene Ontology (GO) analysis for each cultivar, the proteins were grouped, using the corresponding gene code, in unique gene products and subjected to GO analysis through the PANTHER (Protein ANalysis THrough Evolutionary Relationship; version 14.1) system (<http://www.pantherdb.org>) by using the *Triticum aestivum*, *Oryza sativa*, *Brachypodium*, *Hordeum vulgare*, *Arabidopsis thaliana* and *Zea mays* genome annotations as background.

#### Acknowledgments

The authors gratefully acknowledge the Bio-Nanotech Research and Innovation Tower (BRIT; PON project financed by the Italian Ministry for Education, University and Research MIUR). This research was partly supported by “Piano della Ricerca di Ateneo 2016–2018” of the University of Catania, Italy.

#### Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] A. Di Francesco, R. Saletti, V. Cunsolo, B. Svensson, V. Muccilli, P. De Vita, S. Foti, Qualitative proteomic comparison of metabolic and CM-like protein fractions in old and modern wheat Italian genotypes by a shotgun approach, *J. Proteom.* 211 (2020) XXX, <https://doi.org/10.1016/j.jprot.2019.103530>.
- [2] W.J. Hurkman, C.K. Tanaka, Extraction of wheat endosperm proteins for proteome analysis, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 849 (2007) 344–350, <https://doi.org/10.1016/j.jchromb.2006.11.047>.
- [3] R. Saletti, S. Reina, M.G.G. Pittalà, A. Magrì, V. Cunsolo, S. Foti, V. De Pinto, Post-translational modifications of VDAC1 and VDAC2 cysteines from rat liver mitochondria, *Biochim. Biophys. Acta Bioenerg.* 1859 (2018) 806–816, <https://doi.org/10.1016/j.bbabi.2018.06.007>.