Proteome Map of Pea (*Pisum Sativum* L.) Embryos Containing Different Amounts of Residual Chlorophylls

Tatiana Mamontova,[1,2§] Elena Lukasheva,[2§] Gregory Mavropolo-Stolyarenko,[2§] Carsten Proksch,[3] Tatiana Bilova,[1,4] Ahyoung Kim,[1] Vladimir Babakov,[5] Tatiana Grishina,[2] Wolfgang Hoehenwarter,[3] Sergei Medvedev,[4] Galina Smolikova[4*] and Andrej Frolov[1,2*]

[1]Department of Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry

[2]Department of Biochemistry, St. Petersburg State University

[3]Proteome Analytics, Leibniz Institute of Plant Biochemistry

[4]Department of Plant Physiology and Biochemistry, St. Petersburg State University and 5Research Institute of Hygiene, Occupational Pathology and Human Ecology

[§]These authors contributed equally on the manuscript

* Correspondence: afrolov@ipb-halle.de, Tel.: +49 (0) 345 55821350; g.smolikova@spbu.ru, +78123289695

Abstract: Due to low culturing costs and high seed protein contents, legumes represent the main global source of food protein. Pea (*Pisum sativum* L.) is one of the major economically important legume crops, impacting both animal feed and human nutrition. Therefore, the quality of pea seeds needs to be ensured in the context of sustainable crop production and nutritional efficiency. Obviously, changes in seed protein patterns might directly affect both of these aspects. Thus, here we address the pea seed proteome in more detail and provide, to the best of our knowledge, the most comprehensive annotation of the functions and intracellular localization of pea seed proteins. Accordingly, 1938 and 1989 non-redundant proteins were identified in embryos of yellow and green pea seeds, in total. Only 35 and 44 proteins, respectively, could be additionally identified after protamine sulfate precipitation (PSP) potentially indicating the high efficiency of our experimental workflow. In total 981 protein groups could be assigned to 34 functional classes, which were to a large extent differentially represented in yellow and green embryos. Closer analysis of these differences by processing of the data in KEGG and String databases revealed their possible relation to a higher metabolic status and reduced longevity of green embryos.

Keywords: Chlorophylls, LC-MS-based proteomics, pea (*Pisum sativum* L.), proteome functional annotation, proteome map, seeds, seed proteomics

1. Introduction

Legumes represent the most prominent source of food protein, and their importance is increasing with the growing global population [1]. Indeed, these crops are tolerant to environmental stressors, cheap to culture, and rich in seed protein (typically about 25% of fresh seed weight) [2,3]. Among the cultured legumes, pea (*Pisum sativum* L.) is the most widely spread pulse crop in Europe, where it serves as a protein food supplement for monogastric animals [4]. Therefore, quality of pea seeds is important from the aspects of both sustainable crop production and high nutritional efficiency. Obviously, changes in protein composition of seeds might directly affect their agricultural and nutritional value [5].

The first pea seed protein map was reported at the end of the last decade: it was based on the two-dimensional gel electrophoresis (2D-GE) and mass spectrometric (MS) identification of visualized electrophoretic zones (spots) and contained 156 proteins [6]. The majority of the identified polypeptides were storage proteins (convicilins, vicilins, and legumins), which strongly dominate the seed proteome and can serve as seed protein quality markers [7]. Removal of these highly abundant storage proteins by extraction with aqueous (aq.) isopropanol-containing solutions [8], or by

precipitation in presence of aq. 0.01 – 0.1% (w/v) protamine sulfate [9] could increase coverage of the seed proteome. Alternatively, as was shown for soybean (*Glycine max*), seed storage globulins (glycinin and ß-conglycenin) can be effectively removed by 10 mmol/L CaCl₂ [10]. Low-abundant proteins can be selectively enriched by means of the combinational peptide ligand libraries technology [11]. However, despite the efficiency of these techniques, in combination with the gel-based proteomics approach, the numbers of identified proteins never exceed several hundreds.

Because of this, implementation of liquid chromatography (LC)-MS-based strategy is desired to get a deeper insight into the seed proteome [12]. However, as detergents, conventionally used for solubilization of protein isolates [5], dramatically affect efficiency of electrospray ionization (ESI) [13], it is difficult to find a compromise between completeness of protein reconstitution and sensitivity of MS analysis. In this context, the introduction of degradable detergents into proteomic practice helped to overcome this caveat [14]. Thus, commercially available detergents such as RapiGest™ and Anionic Acid-Labile Surfactant II (AALS II) gave a deeper insight in the proteomes of young barley [15] and developing oilseed rape [16] seeds.

When considering the pea seed proteome, it is necessary to remember, that in some cultivars, chlorophylls are not quantitatively destroyed after completion of seed maturation, and the seeds of such plants preserve green color in their mature state [17]. This phenomenon is underlied by deficiency of one or several chlorophyll catabolic enzymes (CCEs), e.g. chlorophyll *b* reductase, 7-hydroxymethyl chlorophyll *a* reductase, Mg²⁺-dechelatase, pheophytinase, pheophorbide *a* oxygenase and reductase of red chlorophyll catabolite (RCC) [17]. Thus, the presence of chlorophylls in *sgr* (*stay-green*) mutants is mostly attributed to the damage of *SGR* genes, prospectively encoding the enzymes involved in chlorophyll degradation and/or disassembly of chlorophyll-protein complexes [18]. Remarkably, green seeds are characterized with higher redox status and amenability to stressors [19]. Adaptation to these metabolic changes might result in essential alterations in seed proteome. These events, potentially affecting seed nutritional properties, have not been addressed so far.

Therefore, here, we provide, to the best of our knowledge, the most complete map of the pea seed proteome using sample pre-fractionation and LC-MS-based shotgun proteomics. We address the differences in embryo proteomes of yellow- and green-seeded pea cultivars and discuss their effects on the function and localization profiles of seed proteins.

2. Results

2.1 Analysis of physiological and biochemical parameters of seed quality

The yellow seeds of the cultivar Millennium and the green ones of the cultivar Gloriosa clearly differed in the contents of photosynthetic pigments. Thus, chlorophylls were detected solely in green seeds, and the contents of carotenoids were approximately 25% higher in these seeds (t-test: $p = 5.41E-09$), in comparison to the yellow ones (Figure 1A and Supplementary information 2). Both pea cultivars demonstrated 100% seed viability, although the seeds of the cultivar Millennium germinated faster (Figure 1B). Moreover, the percentage of normally developed seedlings on day ten was 13% higher for yellow seeds (Figure 1C). Conductivity testing did not reveal differences in electrolyte leakage between yellow and green seeds (t-test: $p = 0.07$, Figure 1D), whereas the green seeds demonstrated statistically significant approximately two- and four-fold higher levels of lipid peroxidation products and H_2O_2 (t-test: $p = 0.0058$ and 0.0002, respectively, Figure 1E and F).

2.2 Protein isolation and tryptic digestion

To ensure efficient extraction of seed proteins, and the maximal coverage of the pea seed proteome, we decided to use phenol-based protein extraction (Figure 2). Thereby, resulting dry protein isolates could be reconstituted in shotgun buffer containing at least 0.15% AALS. Protein determination revealed the extraction yields were in the range of 39.6 –124.1mg/g fresh weight (Table S1-5). Assay precision was determined by SDS-PAGE loading 5 μg of protein (Figure S1-3):
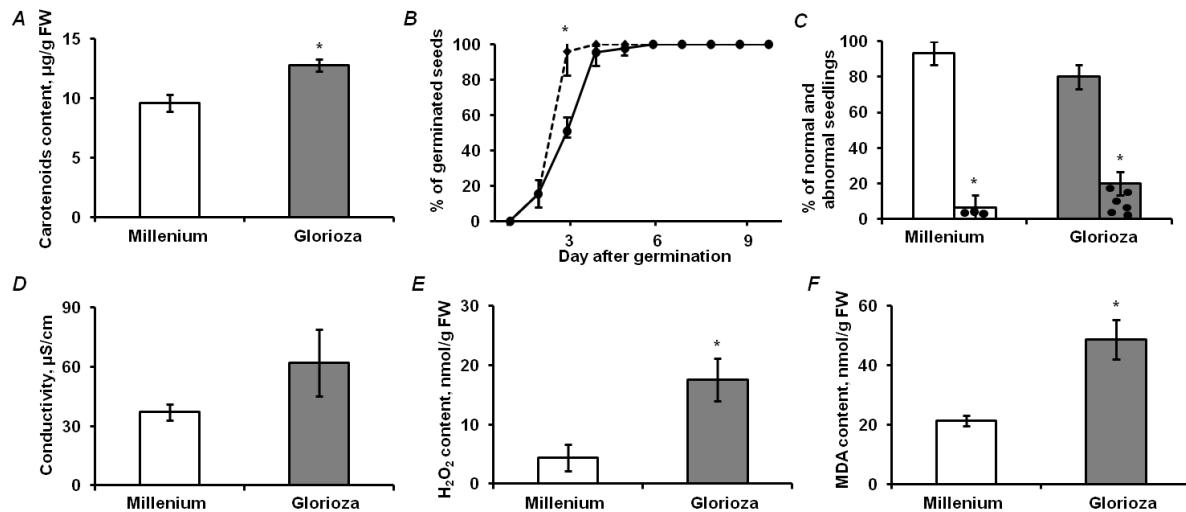
Figure 1 Physiological and biochemical parameters of seed quality, acquired for the mature seeds of cultivar Millennium (solid line, white columns) and Gloriosa (dashed line, grey columns): (**A**) contents of carotenoids, (**B**) kinetics of germination, (**C**) distribution of seedlings by morphology (filled and dotted columns indicate normally and abnormally developed seedlings, respectively), (**D**) electrolyte leakage, expressed as medium conductivity, µS/cm, as well as (**E**) tissue levels of hydrogen peroxide and (**F**) lipid peroxidation products, expressed as malondialdehyde (MDA) equivalents.*Shows statistically significant difference, t-test: $p < 0.05$.

the overall lane densities were $1.4 \times 10^4 \pm 4.5 \times 10^2$ arbitrary units (AU, RSD = 3.06%). The signal patterns observed in the electropherograms were similar between lanes and pea cultivars (Figure S1-3). Tryptic digestion of proteins was considered to be complete, as the bands of major pea storage proteins such as legumin (α- and β- subunits, ~40 kDa and ~20 kDa, correspondingly), vicilin (subunits of ~29 kDa, ~35 kDa and ~47 kDa) and convicilin (subunit of ~71 kDa), could not be detected (Figure S1-4), assuming a staining sensitivity better than 30 ng [20] and a legumin content of at least 80% of total seeds proteins [21].

2.3 Depletion of storage seed proteins by protamine sulfate

To ensure sufficient efficiency of the PSP procedure in the presence of AALS, we diluted the protein samples 10-fold, to arrive at a final AALS concentration of 0.015 % (w/v). Additionally, we applied the highest concentration, tested by the Kim's group (0.07 %), although the authors reported 0.05 %(w/v) protamine sulfate as a sufficient concentration, for efficient depletion of seed storage proteins. Interestingly, the protein recoveries were higher for yellow seed embryos (71.3 – 84.0 mg/g fresh weight, Table S1-6), than for the green ones (56.2 – 61.0 mg/g fresh weight). However, the recoveries after protamine sulfate depletion were slightly higher for green than for yellow seed

Figure 2 Experimental workflow, employed for characterization of the pea seed proteome

embryos (3.7 – 4.2 vs. 2.9 – 3.8 mg/g fresh weight). Accordingly, the depletion efficiency was slightly higher for yellow, than for the green seed embryos (95.4 - 96.0 vs. 92.6 – 94.0 %, Table S1-6). In agreement with this observation, electropherograms of depleted protein extracts were clearly different for yellow and green seeds (Figure S1-5A), although this was not the case for non-depleted samples (Figure S1-3). SDS-PAGE revealed depletion of the major seed storage proteins (Figure S1-3 and S1-5), which, was, however, incomplete (Figure S1-6). Tryptic digestion of depleted samples was complete, i.e. no proteins were detected in the corresponding electropherogram (Figure S1-7).
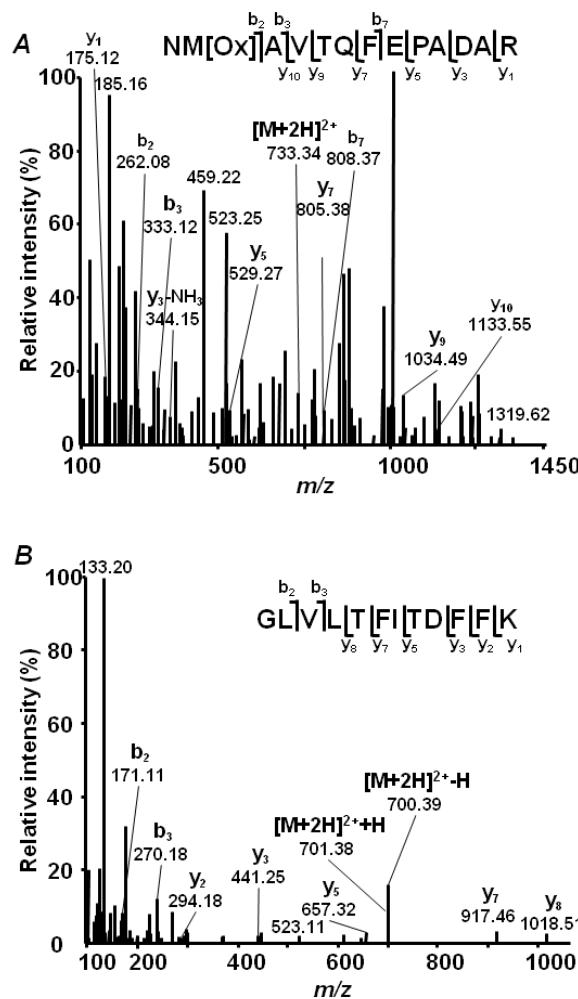
Figure 3 Tandem mass spectra of (**A**) *m/z* 733.3 corresponding to the peptide NM$_{Ox}$AVTQFEPADAR which represents residues 131 – 143 of aminopeptidase (protein accession Lj1g3v1787580.1, *Lotus japonicus*), (**B**) *m/z* 700.89 corresponding to the peptide GLVLTFITDFFK which represents residues 179 – 190 of the eukaryotic translation initiation factor (protein accession G7KRJ1,*Medicago truncatula*), and (**C**)*m/z* 510.79 corresponding to the peptide SVAGEIFGLK which represents residues 170 – 180 of the protein from glutamine synthetase family (protein accessionV7BHN3, *Phaseolus vulgaris*)

2.4 Annotation of pea seed proteins

To verify the applicability of our combined database for annotation of pea seed proteins, we manually interpreted MS/MS spectra surpassing dual FDR thresholds (strict 0.01 and relaxed 0.05), but identified with the lowest XCorrs in both yellow and green seed embryo preparations. This procedure revealed reliable identification of low-scoring peptides, and this was valid for all three proteomes comprising the database: the sequences of corresponding individual peptides could be confirmed by the numbers of b and y fragments, sufficient for their unambiguous identification (Figure 3).

In total 9162 peptides (7923 and 8292 in Millennium and Gloriosa seeds, respectively, Figure 4A, Supplementary information 3) were identified with the FDR of 0.05. On the basis of these identifications, 8769 possible proteins could be annotated (7821 and 8134 in yellow and green seeds, respectively, Figure 4B, 0.05 protein FDR threshold), which represented 2195 non-redundant proteins, or so-called protein groups (1938 and 1989 in Millennium and Gloriosa seeds, respectively, Figure 4C). Precipitation of highly-abundant proteins with 0.07% (w/v) protamine sulfate (PS) resulted in

identification of 2399 and 2286 tryptic peptides in yellow and green seed embryos, respectively (totally 2974, Figure 4A), however only 44 and 85 peptides were unique for PS-treated extracts of yellow and green seeds. These unique peptides, identified specifically in yellow and green seeds, gave access to 11 and 20 protein groups, represented by 24 and 42 proteins, respectively (Figure 4B and C). Further 24 protein groups, represented by 84 proteins were in common to yellow and green seeds, although identified solely in corresponding PS fractions. Removal of the high-abundant proteins was effective, as can be judged by high numbers of proteins and protein groups, observed exclusively in non-depleted samples (Figure 4B

and C). Indeed, 1152 and 1196 proteins were removed by the PS treatment from the extracts of Millennium and Gloriosa seeds (1377 in total), that accounted for 8 and 10 % of the totally identified protein groups.

2.5 Functional annotation of seed proteins and prediction of their cellular localization

Functional annotation resulted in assignment of 981 non-redundant proteins (875 and 900 proteins, isolated from yellow and green seed embryos, respectively) to one or more of 34 functional bins, whereas 1269 polypeptides remained unassigned (Figure 5A, Supplementary information 4a-f). Among the assigned proteins, 926 and 380 could be annotated to functional bins in non-depleted and PSP-depleted fractions, respectively (Figures 5B and C). The bin with the most entries (274 and 275 entries for yellow and green seeds, respectively), was nucleotide metabolism (bin #23), although it included nucleotide binding proteins and ATPases (Supplementary information 4). Bins related to photosynthesis, metal handling, miscellaneous enzymes, RNA, DNA and protein metabolism (# 1, 15, 26 – 29) were also large (Figure 5A). For most of the functional groups, the numbers of proteins assigned to them were similar in yellow and green seed embryos. However, the number of entries in the groups of metal-binding proteins and enzymes of amino acid metabolism were higher in green seed embryos, whereas nucleotide metabolism pathways and miscellaneous enzyme families were more represented in yellow seeds.

The results of localization prediction were in agreement with the observed functional patterns, although only minimal differences between yellow and green seeds were observed (Figure 6, Supplementary information 5a-f).
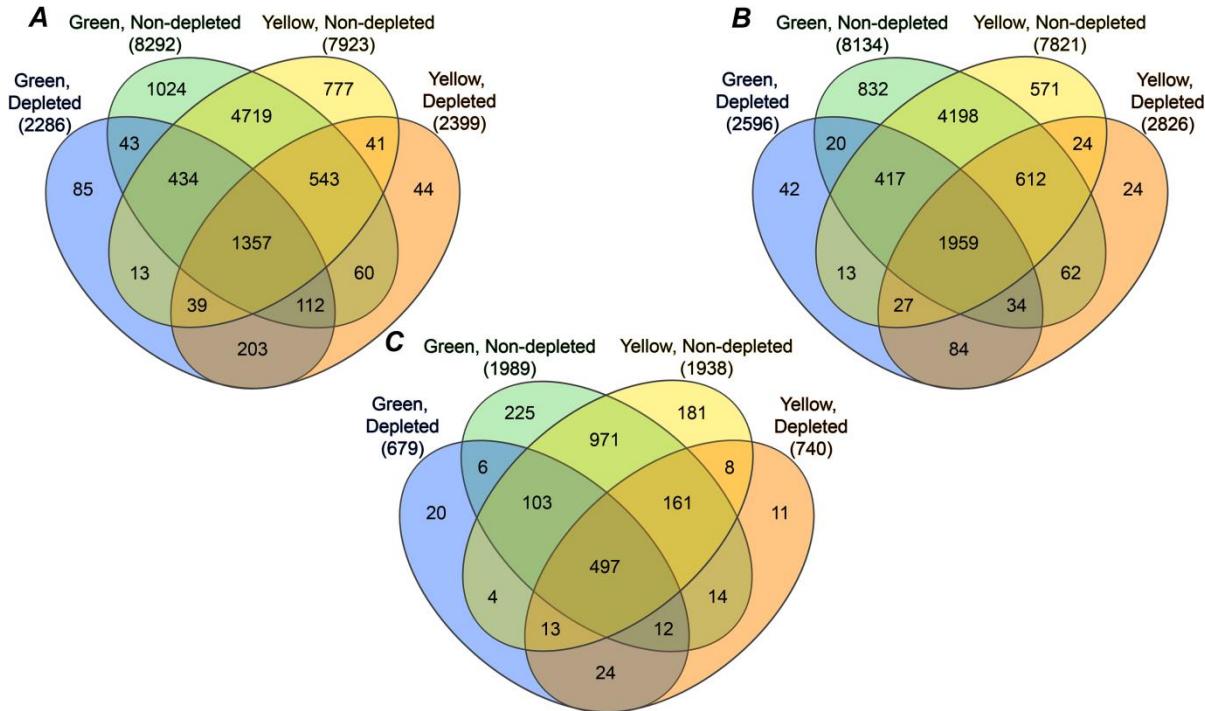
Figure 4 The numbers of tryptic (**A**) peptides, (**B**) proteins, and (**C**) protein groups, identified in yellow-colored (cultivar Millennium) and green-colored (cultivar Gloriosa) embryos with and without depletion of highly abundant proteins with protamine sulfate.
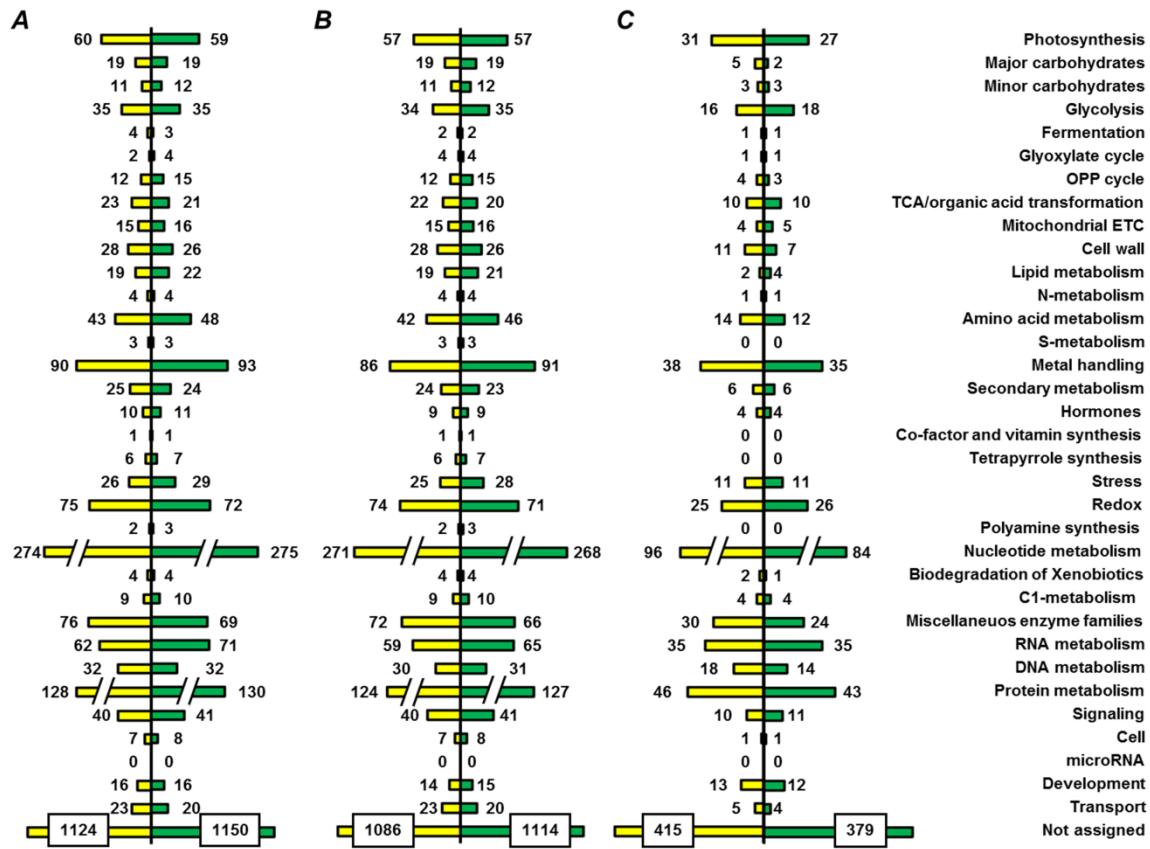
Figure 5 Functional annotation of proteins, identified in embryos of cultivar Millennium (yellow color) and in embryos of cultivar Gloriosa (green color) (**A**) both in non-depleted samples and after treatment with 0.07% (w/v) protamine sulfate, (**B**) only in non-depleted fraction and (**C**) protamine-sulfate supernatant fraction.

Thus, in yellow seed embryos, representation of cytosol and plasma membrane was slightly higher (20.4 and 12.6%), whereas the nuclear fraction accounted for 1.5% less, in comparison with green seeds (Figure 6A). In green seeds, the largest fraction (approximately 46.7%) was represented by nuclear proteins, whereas the majority of the other proteins were localized to cytosol, mitochondria and plasma membrane (19.3, 13.7, and 11.9%, respectively, Figure 6B). Remarkably, depletion of highly-abundant products did not essentially affect this distribution (Figure S1-8).

To address specific metabolic features of yellow and green seeds, we considered the proteins, identified solely in one of the cultivars. Despite the general similarity of protein profiles (Figure 7A, Supplementary information 6), miscellaneous enzyme families were better represented in yellow seeds, whereas the pathways of RNA and amino acid metabolism were more prevalent in green ones. Interestingly, green seeds contained more proteins, which could be localized to the nucleus and endoplasmic reticulum (55 and 2.7%, respectively, Figure 7B), whereas cytosol and cytoskeleton were represented more in yellow seed embryos (15 and 2%, respectively, Figure 7C). Among 251 proteins, unique for green seed embryos (Supplementary information 7), 237 could be annotated by way of homology to *A. thaliana* proteins and analyzed with the String database. Thereby, 9 structural/functional and 7 co-expression protein interaction networks could be annotated (Figure 8).

Finally, we addressed the question, how application of the protamine sulfate precipitation (PSP) affects functional profiles and protein localization. For this, we compared the proteins, identified only in non-depleted fractions (i.e. precipitation in the presence of protamine sulfate) and those, found only after depletion of the highly abundant proteins. As can be seen from Figure 9A, precipitation of highly-abundant proteins (representing mostly nucleotide, protein, amino acid and
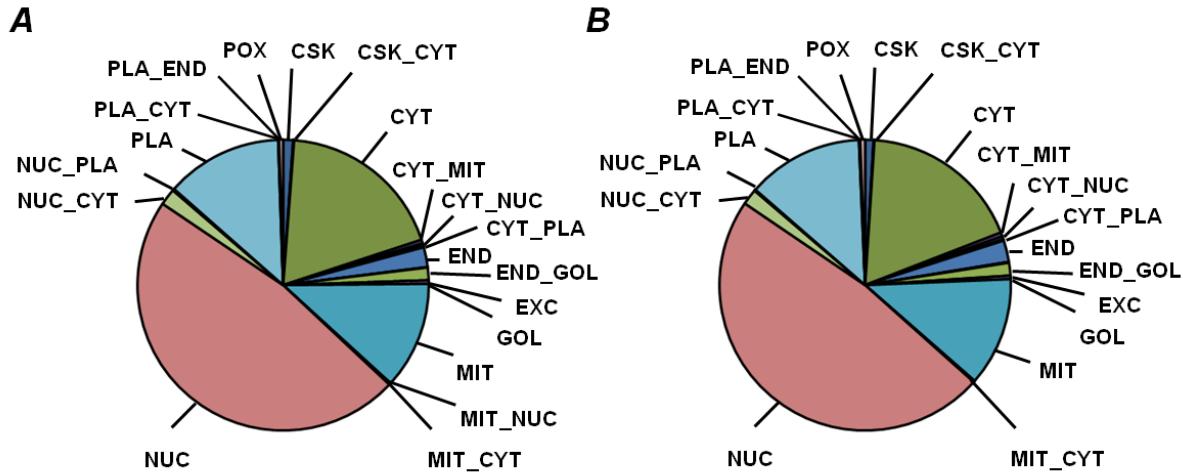
Figure 6 Prediction of sub-cellular localization of proteins, identified (**A**) in yellow embryos of cultivar Millennium and (**B**) in green embryos of cultivar Gloriosa both in non-depleted samples and after treatment with 0.07% (w/v) protamine sulfate. Sub-cellular fractions: CSK, cytoskeleton; CYT, cytoplasm; END, endoplasmic reticulum; EXC, extracellular/secreted; GOL, Golgi apparatus; MIT, mitochondria; NUC, nuclear; PLA, plasma membrane; POX, peroxisome; doubled labels denote corresponding combinations of possible localization

metal-related metabolism, Supplementary information 8) led to the identification of several other proteins annotated to protein, nucleic acid and nucleotide metabolism. Accordingly the supernatants, obtained after supplementation with protamine sulfate, were dominated by nuclear proteins, whereas the depleted proteins exhibited a higher percentage of cytosolic, mitochondrial and plasma membrane polypeptides (Figure 9B, C).

3. Discussion

3.1 Protein extraction and depletion of highly-abundant proteome fraction

Similarly to our earlier study [16], a high variability of protein yields (39.6 – 124.1 mg/g fresh weight, Table S1-5) was observed. This phenomenon can be explained by biochemical heterogeneity of seeds [22,23]. Approximately 50 % higher recoveries, observed for green seeds, were in agreement with a higher number of proteins, identified in this fraction (Figure 4). On the other hand, precise estimates of protein abundance (RSD = 3 %) were in line with our other previous work [24], and thus allowed a comparative, label-free proteomics approach of the heterogeneous seed proteomes [20].
Due to a high abundance of major seed storage proteins [9] and a high amenability of DDA-based shotgun proteomics measurements to a so-called under-sampling effect [25], poor identification rates can be expected for low-abundant proteins. This issue can be expected to be the case when analyzing seed protein digests, as the seed proteome is dominated by several highly abundant protein families – vicilins, convicilins, 11S globulins (legumins), and 2S albumins (PA1 and PA2) [26]. Among the analytical solutions, available for selective removal of storage proteins [27], we selected the PSP method due to its high precision (CV < 12%) and generally, in contrast to most of the other published techniques, quantitative nature [28]. Indeed, in our hands this approach allowed extensive proteome coverage, i.e. only a low number of proteins could be additionally identified after PSP (only 1.7 % of the totally 2195 proteins, Figure 4). Interestingly, PSP yielded mostly regulatory proteins, represented by transcription factors and regulators (GO terms "DNA binding" and "RNA binding"), kinases, transport proteins and enzymes of energy metabolism (Supplementary information 8). The variability of structural domains of these mostly short and polar polypeptides was much lower in PSP-treated samples (Supplementary information 9).
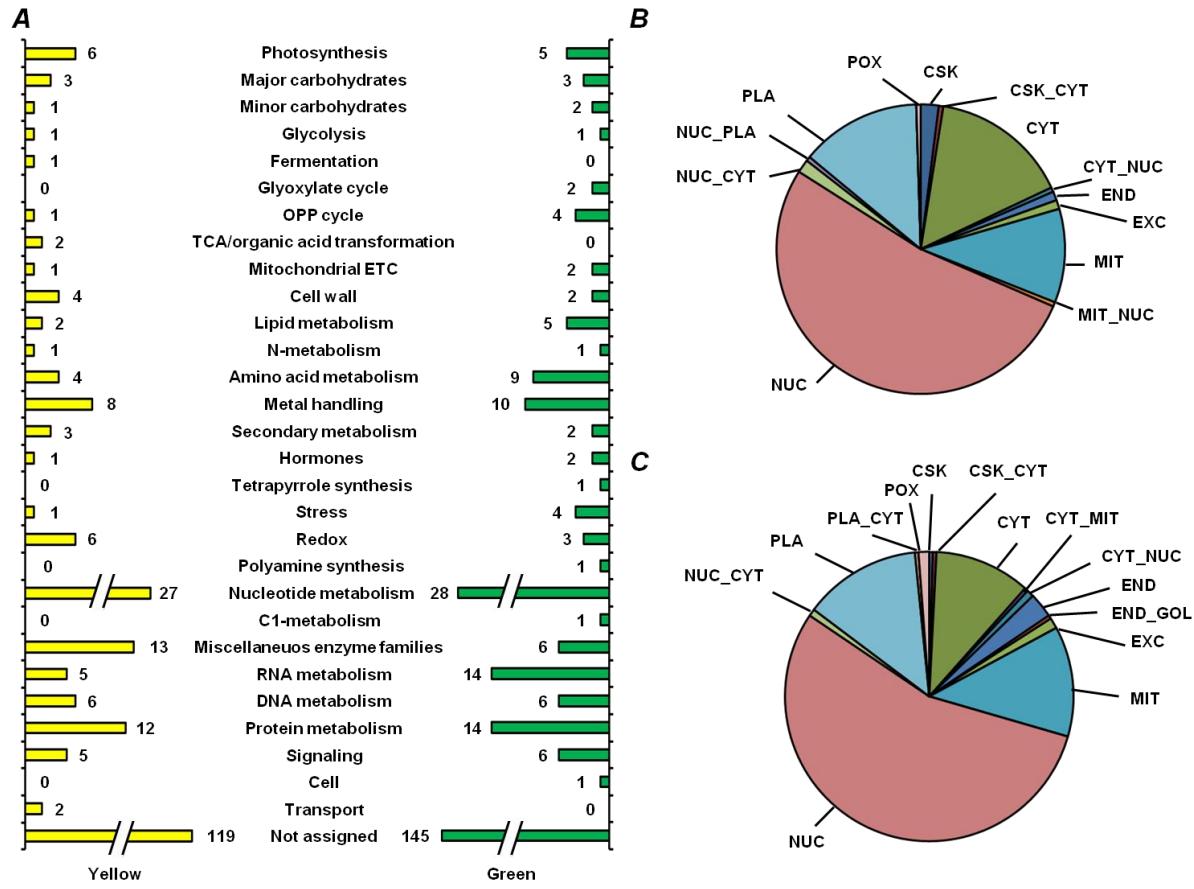
Figure 7 Functional annotation (**A**) and prediction of sub-cellular localization (**B** and **C**) of proteins, unique for yellow embryos of cultivar Millennium (yellow color in panel **A** and **C**) and for green embryos of cultivar Gloriosa (green color in panel **A** and **B**) both in non-depleted samples and after treatment with 0.07% (w/v) protamine sulfate. Sub-cellular fractions: CSK, cytoskeleton; CYT, cytoplasm; END, endoplasmic reticulum; EXC, extracellular/secreted; GOL, Golgi apparatus; MIT, mitochondria; NUC, nuclear; PLA, plasma membrane; POX, peroxisome; doubled labels denote corresponding combinations of possible localization

3.2 Annotation of the pea seed proteome

To date, studies of the legume seed proteome mostly relied on gel-based techniques (typically 2D-GE and MALDI-TOF(/TOF)-MS), yielding up to several hundreds of protein identifications [43, 44]. Sample pre-fractionation [31] and isoelectrofocusing (variation of pH gradients) [29,32] do not dramatically improve the situation. In this context, due to the better resolution of RP-HPLC, an LC-MS-based approach seemed promising [14]. Indeed, the most representative LC-MS-based study to date of Min *et al*, performed with soybean seeds, identified 1626 non-redundant proteins [33], which exceeds the best outcomes from gel-based proteomics surveys by at least two-fold [29]. However it accounts for only a half of the proteins annotated here (Figure 4), although the authors relied on the same pre-fractionation strategy and instrumentation. The possible reasons for this might be (*i*) a longer nanoHPLC column and shallower gradient, (*ii*) a more efficient digestion protocol and (*iii*) a more representative sequence database, used here.

Longer analysis times with shallow gradients might allow better separation of the complex peptide mixtures [34] and a larger number of DDA cycles per run. Our digestion protocol relies on Anionic Acid-Labile Surfactant (AALS, Progenta), which ensures quantitative digestion of all proteins in a sample, and can be completely destroyed and removed, upon proteolysis [35]. Recently, we optimized this procedure for the total protein fraction, obtained by phenolic extraction from various plant species and tissues [16,25].Similarly to the procedure based on the RapiGest detergent (Waters) [15], this protocol is straightforward and does not include steps, at which losses
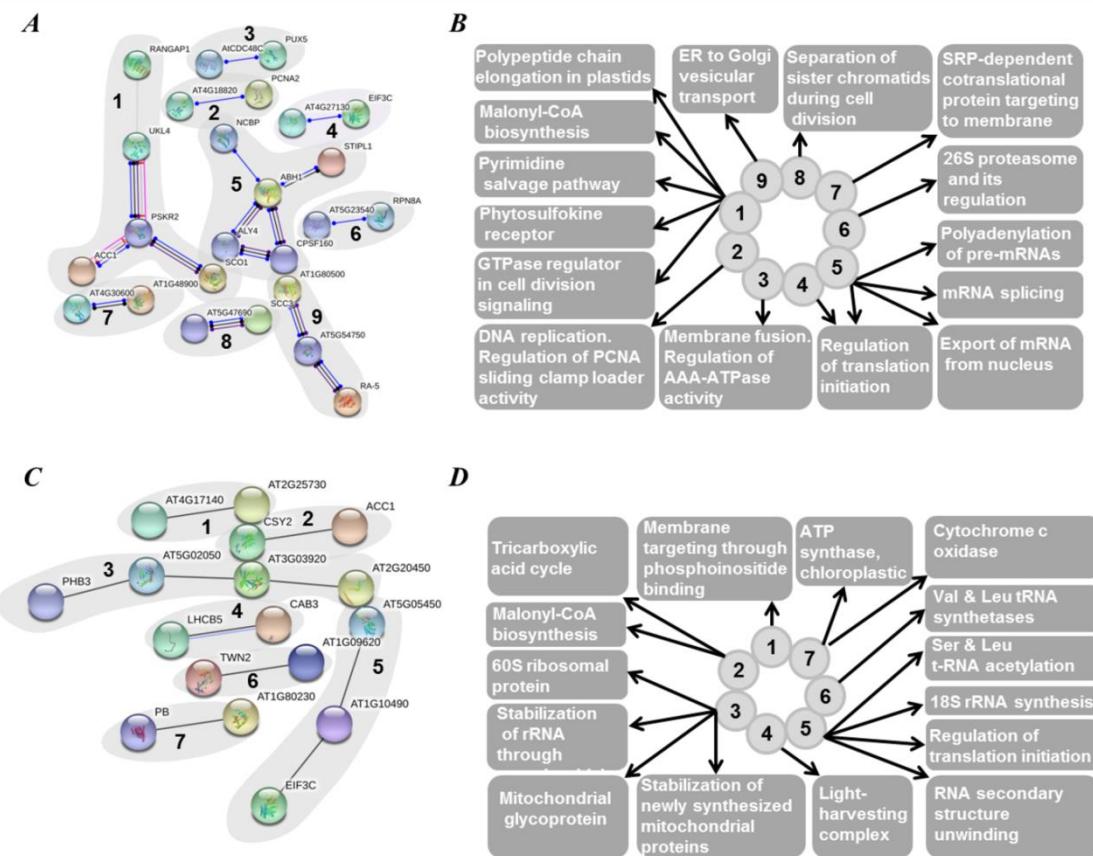
Figure 8 String database analysis (https://string-db.org/) of interaction networks formed by proteins, unique for green embryos (cultivar Gloriosa). The networks (**A**, **C**) and related functional pathways (**B**, **D**) relied on experimentally derived functional and/or structural evidences (**A**, **B**) and co-expression data (**C**, **D**) with interaction score ≥ 0.7 (corresponded to highly-confident entries).Color coding of interactions: black, reaction; blue, binding; light blue, phenotype; green, activation; purple, catalysis; red, inhibition; pink, posttranslational modifications (PTMs); yellow, transcriptional regulation. Filled and empty nodes denote proteins with known and unknown structure, respectively. Detailed description of protein data is provided in Supplementary information 7.

of recovery and/or precision would be expected to occur (e.g. ultra filtration [36] or pre-separation by PAGE [37]).Finally, our sequence database, containing protein entries of three legumes, related to pea, already proved to be a reliable tool for the identification of low-abundant post-translationally modified (glycated) tryptic peptides (known to yield complex fragmentation [38]) in common bean [39]. The database contained exclusively reviewed (i.e. based on transcriptome and proteome data) entries. As reviewed sequence information is currently scarce for *P. sativum*, we assumed this approach to be more appropriate than use of a genomic database. The results of manual interpretation of low-scoring statistically significant PSMs (Figure 3) suggested reliable annotation of peptide sequences.

3.3 Functional annotation of pea seed proteins

Functions of pea seed embryo proteins were successfully annotated by in-house designed processing pipeline. As can be judged by spectral counts (data not shown) for corresponding unique peptides, storage proteins were the most abundant in proteomes of both yellow and green seeds, and their
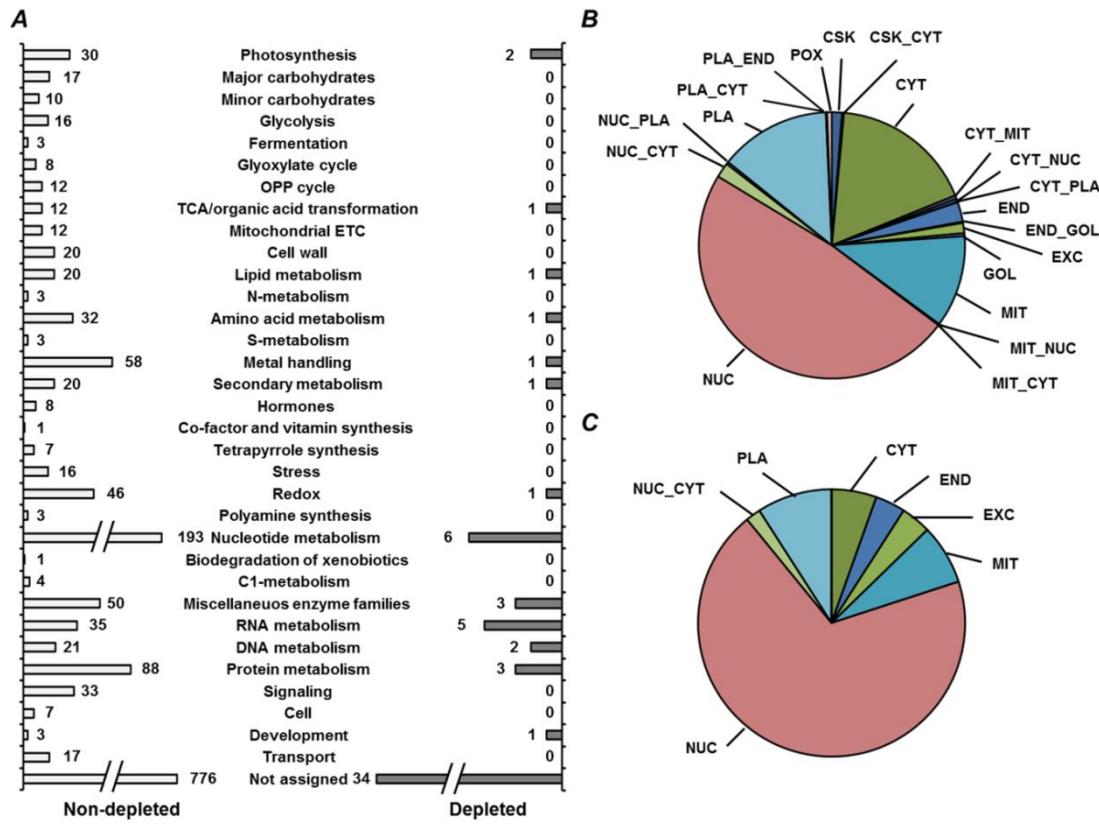
Figure 9 Functional annotation (**A**) and prediction of sub-cellular localization (**B** and **C**) of proteins, identified both in yellow embryos of cultivar Millennium and in green embryos of cultivar Gloriosa in non-depleted samples (white color in panel **A** and **B**) and after treatment with 0.07% (w/v) protamine sulfate (grey color in panel **A** and **C**). Sub-cellular fractions: CSK, cytoskeleton; CYT, cytoplasm; END, endoplasmic reticulum; EXC, extracellular/secreted; GOL, Golgi apparatus; MIT, mitochondria; NUC, nuclear; PLA, plasma membrane; POX, peroxisome; doubled labels denote corresponding combinations of possible localization

pattern – legumins, convicilins, vicilins and glycinins corresponded well with the observation of Bourgeois *et al* [6] (data not shown). The other major functional classes were represented by the polypeptides involved in metabolism of proteins and RNA, photosynthesis, metal handling, redox metabolism and general metabolism enzyme families (Figure 5). Based on the number of functionally assigned proteins (Supplementary information 4), representation of protein metabolism in pea seeds was comparable to soybeans [37,40]. However, the metal handling, RNA metabolism, and redox groups were much more abundant here in comparison to those studies. This can be explained by deeper coverage of the seed proteome, (3-fold compared to Han *et al* [37]), although the incomplete functional assignment in our study, makes the difference less striking.

3.4 Features of embryonic proteome, related to seed vigor: impact of residual chlorophylls

As pea seeds belong to the orthodox type, and acquire desiccation tolerance during development [41], their ability to sustain a prolonged dehydration is the most important feature underlying their high vigor (i.e. the properties determining the potential for rapid, uniform emergence, and development of normal seedlings [42]). During desiccation, stability of cellular structures is secured by accumulation of protective proteins and quenching of over-produced reactive oxygen species (ROS) [43]. Here, the green seeds showed distinctly lower vigor in comparison to the yellow ones, as can be seen from delayed germination and higher percentage of abnormal seedlings observed on the third day (t-test: p = 0.02, Figure 1B and C). To a large extent, it can be explained by a significantly lower status of their antioxidant system (Figure 1E and F), whereas residual chlorophylls could underlie accelerated formation of ROS [44,45]. Thereby, slightly increased contents of carotenoids

(Figure 1A) didn't provide sufficient to quench ROS, which could lead to slightly compromised stability of membrane structures (Figure 1D).

To address the effects prospectively related to seed residual chlorophylls at the molecular level, we considered the proteins potentially affecting desiccation tolerance, vigor and longevity. Among them, late embryogenesis abundant (LEA) proteins preserving cell integrity under desiccation stress [46], were one of the most highly represented groups, accounting for 9 proteins (Supplementary information 3 and 4). Another important feature underlying high vigor is pre-expressed protein synthesis machinery, which allows fast and concerted germination of seeds and development of seedlings [43]. Not less importantly, rapid degradation of storage proteins is a pre-requisite for appropriate mobilization of plant resources for growth of a new plant organism [47]. Also, damaged (e.g. oxidized, glycated, lipoxidized or miss folded) polypeptides need to be degraded to prevent deleterious effects within the cell [48,49]. In agreement with previous reports [33], proteins involved in protein metabolism, were represented with t-RNA ligases, translation initiation, elongation factors, chaperones, and degradation enzymes.

Some enzymes related to protein metabolism showed differential expression in yellow- and green-seeded cultivars. Thus, totally 33 non-redundant t-RNA ligases were identified in yellow pea seed embryos, while six unique proteins were additionally annotated in green ones. Translation initiation factors demonstrated similar differences (Supplementary information 4a and 6). As a decrease in abundance of translation factors is known to accompany age-related loss of seed vigor [50,51], the observed expression profiles might indicate a higher amenability of green seed embryos to such a loss during a prolonged dormancy period. Protein folding was represented by more than 20 entries (HSP70 and 90, GroEL, T-complex protein TCP-1/cpn60, Supplementary information 4a, bin 23). Indeed, HSP70 and GroEL were responsive to osmopriming in Arabidopsis and alfalfa seeds, respectively [52,53], whereas abundance of TCP-1 was decreased in aged alfalfa seeds [54]. Remarkably, several peptidases and DnaJ heat shock protein were identified only in Millennium seeds (Supplementary information 6), facilitating faster degradation of damaged polypeptides and faster mobilization of storage proteins. Interestingly, pea seeds contained acylamino-acid-releasing enzyme-like protein, involved in the degradation of oxidized and glycated proteins [55]. Thus, together with the glyoxalase system and ribulosamine/erythrulosamine 3-kinase it might contribute to the plant anti-glycative machinery.

One of the critical factors affecting vigor, longevity and, in general, viability of dormant and germinating seeds is the abundance of ROS in seed tissues [56,57]. In agreement with this, a rich pattern of redox proteins (thioredoxin reductase, glutathione peroxidase, monodehydroascorbate reductase, superoxide dismutase, dehydroascorbate reductase, and catalase), earlier identified in other species [30,37,58], was confirmed here (Supplementary information 4). Yellow seed embryos contained eight unique oxidoreductases, that might indicate higher levels of oxidative catabolism, higher capacity of electron transfer chains [46], and could potentially increase vigor of yellow seeds.

Deeper analysis of metabolic pathways, potentially involved in seed longevity [33,59] (Supplementary information 7) revealed S-adenosylmethionine (SAM) synthase 2 and glycolytic enzyme hexokinase 1 and E1 alpha subunit of the pyruvate dehydrogenase complex. As was earlier shown in priming and ageing experiments [50], SAM synthase is critical for seed vigor. Conversely, enzymes of primary metabolism were shown to be affected by controlled deterioration (accelerated ageing) of seeds, i.e. age-related loss of vigor [60,61].In this case, the presence of additional isoforms of hexokinase and pyruvate dehydrogenase might not only affect the rate of glycolysis, but also increase production of Ac-KoA and its involvement in the TCA cycle. Remarkably, yellow seeds express three unique aldehyde dehydrogenases   - enzymes, reducing reactive carbonyl products of lipid peroxidation and glycation to corresponding alcohols, thereby protecting proteins from oxidative damage [62].

To detect protein-protein interaction networks specific to green seed embryos we performed String database analysis with the proteins found solely in Gloriosa seeds (Figure 8). The largest functional/structural cluster (# 1, Figure 8A) was represented by enzymes, involved in nucleotide metabolism and nucleotide-dependent regulatory pathways, as well as protein biosynthesis in

chloroplasts and malonyl-CoA biosynthesis (Figure 8B). These functional relations can be explained by simultaneous synthesis of lipids and integral proteins of thylakoid membranes, triggered by activation of phytosulfokine receptors and mediated by calmoduline-related signaling [63].The second major cluster (# 5) comprised proteins related to m-RNA processing and initiation of translation. As can be seen from Figure 8A, it is closely related to the previous one, suggesting an impact on activation of gene expression during growth [64]. Analysis of the major co-expression clusters (Figure 8C, clusters 3 and 5) was in agreement with these results, and confirmed activation of protein synthesis (mostly translation and RNA processing directly related to it, Figure 8D) as one of the main features of green-colored seeds (Figure 8A and Supplementary information 7). Further, the co-expression clusters 4 and 7 confirmed the functional annotation data, indicating high representation of photosynthesis-related proteins in the seed proteome (Figure 5).

4. Materials and methods

4.1 Reagents and plant material

Unless stated otherwise, materials were obtained from the following manufacturers. Carl Roth GmbH & Co (Karlsruhe, Germany): acetonitrile (≥99,95 %, LC-MS grade), ethanol(≥99,8%), sodium dodecyl sulfate (SDS) (>99%), tris-(2-carboxyethyl)-phosphine hydrochloride (TCEP, ≥98%); PanReac AppliChem (Darmstadt, Germany): acrylamide (2K Standard Grade), glycerol (ACS grade); AMRESCO LLC (Fountain Parkway Solon, USA): ammonium persulfate (ACS grade), glycine (biotechnology grade), N,N′-methylene-bis-acrylamide (ultra pure grade), tris(hydroxymethyl)aminomethane (tris, ultra pure grade), urea (ultra pure grade), ammonium bicarbonate (puriss, p.a.); Bioanalytical Technologies 3M Company (St. Paul, USA): Empore™ solid phase octodecyl extraction discs; Component-Reactiv (Moscow, Russia): phosphoric acid (p.a.); Reachem (Moscow, Russia): hydrochloric acid (p.a.), isopropanol (reagent grade), potassium chloride (reagent grade); SERVA Electrophoresis GmbH (Heidelberg, Germany): Coomassie Brilliant Blue G-250, 2-mercaptoethanol (research grade), trypsin NB (sequencing grade, modified from porcine pancreas); Thermo Fisher Scientific (Waltham, USA): Pierce™ Unstained Protein Molecular Weight Marker #26610 (14.4–116.0 kDa), PageRuller™ Plus Prestained Protein Ladder #26620 (10–250 kDa); Dichrom GmbH (Marl, Germany): Progenta™ anionic acid labile surfactant II (AALS II) and adaptors for stage-tips; Vekton (St. Petersburg, Russia): acetonitrile (HPLC grade), sucrose (ACS grade), conc. HCl (puriss). All other chemicals were purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany). Water was purified in house (resistance 5-15 mΩ/cm) on a water conditioning and purification systems Elix 3 UV (Millipore, Moscow, Russia) or Millipore Milli-Q Gradient A10 system (resistance 5-15 mΩ/cm, Merck Millipore, Darmstadt, Germany).

Pea seeds of the yellow-seeded cultivar Millennium (Figure S1-1A-D) were obtained from the Research and Practical Center of National Academy of Science of the Republic of Belarus for Arable Farming (Zhodino, Belarus, harvested in the year 2015). The seeds of the green-seeded cultivar Gloriosa (Figure S1-1E-H) were provided by the Institute of Vegetable Growing, National Academy of Science (Minsk, Belarus, harvested in the year 2015).

4.2 Analysis of physiological and biochemical parameters of seed quality

The seeds were germinated on filter paper in a thermostat at 22° C. To address the kinetics of germination, the numbers of germinated seeds were monitored for ten days on a daily basis, before the seeds were classified in (i) non-germinating, and those producing (ii) normal and (iii) abnormal seedlings upon germination, as defined by the International Seed Testing Agency (ISTA) [65]. Electrical conductivity test relied on the method of Matthews et al [66]. For this, seeds (n = 5) were incubated in 25 mL of distilled water for 2 h at 22° C. The conductivity of the aqueous medium was determined with a conductometer HI 8733 (HANNA Instruments Deutschland GmbH, Vöhringen, Germany) before and after the incubation. Determination of photosynthetic pigments (n = 3) relied on the method of Lichtenthaler and Wellburn [67] (for details see Supplementary information 1, Protocol S1-1). Hydrogen peroxide was quantified by the method of Bilova et al [68], whereas the

contents of lipid peroxidation products were determined as malondialdehyde (MDA) equivalents, as described by Frolov and co-workers [24] (see Protocols S1-2 and S1-3, respectively).

4.3 Protein isolation

Pea embryos (10 per biological replicate, n = 3) were frozen in liquid nitrogen and ground in a Mixer Mill MM 400 ball mill with a Ø 20 mm stainless steel ball (Retsch, Haan, Germany) at a vibration frequency of 30 Hz for 2 x 1 min. The resulting ground material (approximately 50 mg per replicate) was placed in 2 mL safe-lock polypropylene tubes and stored at -80°C prior to protein extraction, which relied on the method of Frolov and co-workers [16] with some modifications. Briefly, the tubes with plant material were transferred on ice, and one min later supplemented with 700 µL of cold (4 °C) phenol extraction buffer, containing 0.7 mol/L sucrose, 0.1 mol/L KCl, 5 mmol/L ethylenediaminetetraacetic acid (EDTA), 2% (v/v) mercaptoethanol and 1 mmol/L phenylmethylsulfonyl fluoride (PMSF) in 0.5 mol/L tris-HCl buffer (pH 7.5). The suspensions were vortexed for 30 s, before 700 µL of cold phenol (4 °C) saturated with 0.5 mol/L tris-HCl buffer (pH 7.5) were added. After further mixing for 30 s, the samples were shaken for 30 min at 900 rpm (4 °C) and centrifuged (5000 g, 30 min, 4 °C). Afterwards, the phenolic (upper) phases were transferred to new 1.5 mL polypropylene tubes, and washed two times with equal volumes of the phenol extraction buffer (after each buffer addition - vortexing 30 s, shaking for 30 min at 900 rpm at 4 °C and centrifugation at 5000 g for 15 min at 4 °C). Then, the supernatants were transferred to 1.5 mL polypropylene tubes, and proteins were precipitated by addition of a five-fold volume of ice-cold 0.1 mol/L ammonium acetate in methanol, followed by storage overnight at -20 °C. Next morning, the protein fraction was collected by centrifugation (10 min, 5000 g, 4 °C), and the supernatants were discarded. The pellets were washed twice with two volumes of methanol (relative to the volume of the phenol phase), and twice – with the same volume of acetone (both at 4°C). Each time, the samples were centrifuged (5000 g, 10 min, 4° C) after re-suspension. Finally, the cleaned pellets were dried under air flow in a fume hood for 1 h, reconstituted in 100 µL of shotgun buffer (8 mol/L urea, 2 mol/L thiourea, 0.15% AALS II in 100 mmol/L tris-HCl, pH 7.5), and protein contents were determined by 2-D Quant Kit (GE Healthcare, Taufkirchen, Germany) according to Matamoros and co-workers [39] or by the Bradford method as described by Frolov and co-workers [69]. The precision of the assay was verified by SDS-PAGE according Greifenhagen et al [70] with minor modifications (for details see Protocol S1-4).

4.4 Depletion of seed storage proteins by protamine sulfate and tryptic digestion

Depletion of seed storage proteins relied on the protamine sulfate precipitation procedure (PSP) of Kim et al [9] with some modifications. In detail, 95 µL of the total protein fractions, isolated from green and yellow seeds (n = 3) as described above, were supplemented with 855 µL of 0.077% (w/v) protamine sulfate in 20 mmol/L MgCl2/0.5 mol/L Tris-HCl (pH 8.3) to yield a final protamine sulfate concentration of 0.07% (w/v). After centrifugation (12000 g, 10 min, 4° C), supernatants were transferred to 5-mL polypropylene tubes and supplemented with four volumes of 12.5% trichloroacetic acid (TCA) in cold (-20° C) acetone and left at -20° C overnight. Next morning, the pellets were washed three times with 1 mL of cold acetone by sequential re-suspending and centrifugation (12000 g, 10 min, 4° C), and clean pellets were dried in a fume hood for 1 h. The dried pellets were reconstituted in 100 µL of shotgun buffer containing 0.15% AALS, and the protein contents were determined by 2-D Quant Kit without further dilution. The digestion procedure relied on the protocol of Frolov and co-workers [24] with minor modifications (See Protocol S1-5 for details).

4.5 NanoHPLC-ESI-Q-Orbitrap analysis

Dried tryptic digests were reconstituted in 100 µL of aq. acetonitrile (final concentration of 3% v/v) containing 0.1% (v/v) aq. formic acid, and 1 µL (250ng) of peptide mixture was loaded onto an Acclaim PepMap 100 C18 trap column (75µm x 20mm, nano-viper, 3µm particle size). After trapping, the peptides were separated on a PepMap™ RSL C18 column (75µm x 50 cm, 2 µm particle size)

using an EASY-nLC 1000HPLC system coupled on-line to a Q-Exactive Plus mass spectrometer via an EASY-spray nano ion source (all Thermo Fisher Scientific, Bremen, Germany). The details of the chromatographic method are summarized in Supplementary information 1 (Table S1-1). The nano-LC-Orbitrap-MS analysis relied on data-dependent acquisition (DDA) experiments performed in the positive ion mode, comprising a survey Orbitrap-MS scan and dependent MS/MS scans for the most abundant signals in the following 5 s (at certain tR) with charge states ranging from 2 to 6. The mass spectrometer settings and DDA parameters are summarized in Table S1-2. The database search relied on Proteome Discoverer 2.2 software (Thermo Fisher Scientific, Bremen, Germany), SEQUEST search engine and a combined sequence database comprising proteomes of *Medicago truncatula* Gaertn, *Lotus japonicus* (Regel) K. Larsen and *Phaseolus vulgaris* L [39] with addition of protein sequences, prospectively related to seed longevity and color (see Table S1-3 for details). The redundancy of the database was eliminated by the CD-HIT algorithm [71] with the sequence identity cut-off set to 1. The database search parameters are summarized in Table S1-4.

4.6 Data analysis and post-processing

Functional annotation of pea proteins was done with Visual Basic for Applications (VBA). The corresponding script (Figure S1-2) was run within Excel instance and managed automatic extraction, appendment and rearrangement of tabular data, organized in spread sheets. Further, it established access to a laboratory web server via a simplified HTTP protocol. The server, in turn, was run under Apache 2.4, used as a CGI front end for a set of bash scripts. These scripts managed previously cached data, performed actual queries to external public databases or ran locally installed programs. After screening against UniProt [72] or KEGG [73] databases, amino acid sequences, gene ontology (GO) identifiers and KEGG Orthology (KO) terms with text annotations were extracted as text files [74]. Annotation relied on the functional bins used by the MapMan software suite [75]. Cellular localization was predicted by the NGLOC algorithm [76]. Finally, all annotations were fitted back into the original Excel spreadsheet and clustered by localization terms or by top MapMan hierarchy levels, before the data was manually inspected, corrected or further automatically processed.

Analysis of protein networks relied on a search against the Arabidopsis thaliana proteome in the String database (for details see Protocol S1-6). Only highly-confident protein entries with interaction scores ≥ 0.7 were exported for building protein interaction maps based on (i) experimentally confirmed evidences and (ii) co-expression. Venn diagrams were constructed using the web-based tool InteractiVenn [77] based on the information given in Proteome Discoverer 2.2 output in columns "Found in Samples".

5. Conclusions

The quality of plant seeds is a pre-requisite for sustainable agriculture and food production. As legume crops represent the main source of food protein worldwide, the quality of their seeds requires special attention. Using comprehensive methods of LC-MS based proteomics we succeeded to extensively cover the pea seed proteome and to assign functions to individual non-redundant proteins. This systematic approach allowed the identification of characteristic metabolic features of green-coloured seeds in comparison to yellow-coloured ones. Thereby, alterations in the proteome of green seeds, prospectively related to their reduced vigour and longevity, were revealed.

Author Contributions: Conceptualization, Sergei Medvedev, Galina Smolikova and Andrej Frolov; Data curation, Gregory Mavropolo-Stolyarenko, Tatiana Bilova and Ahyoung Kim; Formal analysis, Tatiana Mamontova, Elena Lukasheva, Gregory Mavropolo-Stolyarenko, Tatiana Bilova and Ahyoung Kim; Funding acquisition, Sergei Medvedev; Investigation, Tatiana Mamontova, Elena Lukasheva, Carsten Proksch, Tatiana Bilova and Tatiana Grishina; Methodology, Galina Smolikova and Andrej Frolov; Project administration, Galina Smolikova; Resources, Tatiana Grishina; Software, Gregory Mavropolo-Stolyarenko; Supervision, Andrej Frolov; Validation, Andrej Frolov; Visualization, Tatiana Mamontova and Elena Lukasheva; Writing – original

draft, Tatiana Mamontova, Elena Lukasheva, Vladimir Babakov and Galina Smolikova; Writing – review & editing, Vladimir Babakov, Wolfgang Hoehenwarter, Sergei Medvedev, Galina Smolikova and Andrej Frolov.

Funding: This research was funded by the Russian Science Foundation (project number 16-16-00026).

Acknowledgments: The authors are grateful to Veronika Chantseva, Olga Shiroglazova and Svetlana Milrud (St. Petersburg State University) for technical assistance. The St. Petersburg St. University Research Resource Center for Molecular and Cell Technologies is acknowledged for technical support.

Abbreviations

2D-GE	two-dimensional gel electrophoresis
AALS II	Anionic Acid-Labile Surfactant II
CCEs	chlorophyll catabolic enzymes
DDA	data dependent acquisition
ESI	electrospray ionization
FDR	false discovery rate
ISTA	International Seed Testing Agency
LEA	late embryogenesis abundant
LC	liquid chromatography
MALDI	matrix-assisted laser desorption/ionization
MDA	malondialdehyde
MS	mass spectrometry
PS	protamine sulfate
PSP	protamine sulfate precipitation
PSMs	peptide spectrum matches
RCC	red chlorophyll catabolite
ROS	reactive oxygen species
RP-HPLC	reverse phase high performance liquid chromatography
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGR	stay green
SPE	solid phase extraction
TOF	time of flight

References

1. Singh, N. Pulses: an overview. *J Food Sci Technol* **2017**, *54*, 853–857, doi:10.1007/s13197-017-2537-4.

2. Iqbal, A.; Khalil, I.A.; Ateeq, N.; Sayyar Khan, M. Nutritional quality of important food legumes. *Food Chemistry* **2006**, *97*, 331–335, doi:10.1016/j.foodchem.2005.05.011.

3. Babar, M.M.; Zaidi, N.S.; Azooz, M.M.; Kazi, A.G. Genetic and molecular responses of legumes in a changing environment. In *Legumes under Environmental Stress*; Wiley-Blackwell, 2015; pp. 199–214 ISBN 978-1-118-91709-1.

4. Casey, R.; Domoney, C.; Smith *Biochemistry and molecular biology of seed products.*; 1993; ISBN 978-0-85198-863-4.

5. Andrej Frolov; Tatiana Mamontova; Christian Ihling; Elena Lukasheva; Mikhail Bankin; Veronika Chantseva; Maria Vikhnina; Alena Soboleva; Julia Shumilina; Gregory Mavropolo-Stolyarenko; Tatiana Grishina; Natalia Osmolovskaya; Vladimir Zhukov; Wolfgang Hoehenwarter; Andrea Sinz; Igor Tikhononovich; Ludger Wessjohann; Tatiana Bilova; Galina Smolikova; Sergei Medvedev Mining seed proteome: from protein dynamics to modification profiles. *BioComm* **2018**, *63*, doi:10.21638/spbu03.2018.106.

6. Bourgeois, M.; Jacquin, F.; Savois, V.; Sommerer, N.; Labas, V.; Henry, C.; Burstin, J. Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics* **2009**, *9*, 254–271, doi:10.1002/pmic.200700903.

7. Bourgeois, M.; Jacquin, F.; Cassecuelle, F.; Savois, V.; Belghazi, M.; Aubert, G.; Quillien, L.; Huart, M.; Marget, P.; Burstin, J. A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics* **2011**, *11*, 1581–1594, doi:10.1002/pmic.201000687.

8.　Natarajan, S.S.; Krishnan, H.B.; Lakshman, S.; Garrett, W.M. An efficient extraction method to enhance analysis of low abundant proteins from soybean seed. *Anal. Biochem.* **2009**, *394*, 259–268, doi:10.1016/j.ab.2009.07.048.

9.　Kim, Y.J.; Wang, Y.; Gupta, R.; Kim, S.W.; Min, C.W.; Kim, Y.C.; Park, K.H.; Agrawal, G.K.; Rakwal, R.; Choung, M.-G.; Kang, K.Y.; Kim, S.T. Protamine sulfate precipitation method depletes abundant plant seed-storage proteins: A case study on legume plants. *Proteomics* **2015**, *15*, 1760–1764, doi:10.1002/pmic.201400488.

10.　Krishnan, H.B.; Oehrle, N.W.; Natarajan, S.S. A rapid and simple procedure for the depletion of abundant storage proteins from legume seeds to advance proteome analysis: a case study using Glycine max. *Proteomics* **2009**, *9*, 3174–3188, doi:10.1002/pmic.200800875.

11.　Boschetti, E.; Righetti, P.G. Plant proteomics methods to reach low-abundance proteins. *Methods Mol. Biol.* **2014**, *1072*, 111–129, doi:10.1007/978-1-62703-631-3_9.

12.　Cerna, H.; Černý, M.; Habánová, H.; Šafářová, D.; Abushamsiya, K.; Navrátil, M.; Brzobohatý, B. Proteomics offers insight to the mechanism behind Pisum sativum L. response to pea seed-borne mosaic virus (PSbMV). *J Proteomics* **2017**, *153*, 78–88, doi:10.1016/j.jprot.2016.05.018.

13.　Vissers, J.P.; Chervet, J.P.; Salzmann, J.P. Sodium dodecyl sulphate removal from tryptic digest samples for on-line capillary liquid chromatography/electrospray mass spectrometry. *J Mass Spectrom* **1996**, *31*, 1021–1027, doi:10.1002/(SICI)1096-9888(199609)31:9<1021::AID-JMS384>3.0.CO;2-G.

14.　Soboleva, A.; Schmidt, R.; Vikhnina, M.; Grishina, T.; Frolov, A. Maillard Proteomics: Opening New Pages. *Int J Mol Sci* **2017**, *18*, doi:10.3390/ijms18122677.

15.　Kaspar-Schoenefeld, S.; Merx, K.; Jozefowicz, A.M.; Hartmann, A.; Seiffert, U.; Weschke, W.; Matros, A.; Mock, H.-P. Label-free proteome profiling reveals developmental-dependent patterns in young barley grains. *J Proteomics* **2016**, *143*, 106–121, doi:10.1016/j.jprot.2016.04.007.

16.　Frolov, A.; Didio, A.; Ihling, C.; Chantzeva, V.; Grishina, T.; Hoehenwarter, W.; Sinz, A.; Smolikova, G.; Bilova, T.; Medvedev, S. The effect of simulated microgravity on the Brassica napus seedling proteome. *Functional Plant Biol.* **2018**, *45*, 440–452, doi:10.1071/FP16378.

17.　Smolikova, G.; Dolgikh, E.; Vikhnina, M.; Frolov, A.; Medvedev, S. Genetic and Hormonal Regulation of Chlorophyll Degradation during Maturation of Seeds with Green Embryos. *Int J Mol Sci* **2017**, *18*, doi:10.3390/ijms18091993.

18.　Hörtensteiner, S. Stay-green regulates chlorophyll and chlorophyll-binding protein degradation during senescence. *Trends Plant Sci.* **2009**, *14*, 155–162, doi:10.1016/j.tplants.2009.01.002.

19.　Smolikova, G.N.; Laman, N.A.; Boriskevich, O.V. Role of chlorophylls and carotenoids in seed tolerance to abiotic stressors. *Russ J Plant Physiol* **2011**, *58*, 965, doi:10.1134/S1021443711060161.

20.　Frolov, A.; Blüher, M.; Hoffmann, R. Glycation sites of human plasma proteins are affected to different extents by hyperglycemic conditions in type 2 diabetes mellitus. *Anal Bioanal Chem* **2014**, *406*, 5755–5763, doi:10.1007/s00216-014-8018-y.

21.　Barac, M.; Cabrilo, S.; Pesic, M.; Stanojevic, S.; Zilic, S.; Macej, O.; Ristic, N. Profile and functional properties of seed proteins from six pea (Pisum sativum) genotypes. *Int J Mol Sci* **2010**, *11*, 4973–4990, doi:10.3390/ijms11124973.

22.　Gallardo, K.; Thompson, R.; Burstin, J. Reserve accumulation in legume seeds. *C. R. Biol.* **2008**, *331*, 755–762, doi:10.1016/j.crvi.2008.07.017.

23.　Thompson, R.; Burstin, J.; Gallardo, K. Post-genomics studies of developmental processes in legume seeds. *Plant Physiol.* **2009**, *151*, 1023–1029, doi:10.1104/pp.109.143966.

24.　Frolov, A.; Bilova, T.; Paudel, G.; Berger, R.; Balcke, G.U.; Birkemeyer, C.; Wessjohann, L.A. Early responses of mature Arabidopsis thaliana plants to reduced water potential in the agar-based polyethylene glycol infusion drought model. *J. Plant Physiol.* **2017**, *208*, 70–83, doi:10.1016/j.jplph.2016.09.013.

25.　Paudel, G.; Bilova, T.; Schmidt, R.; Greifenhagen, U.; Berger, R.; Tarakhovskaya, E.; Stöckhardt, S.; Balcke, G.U.; Humbeck, K.; Brandt, W.; Sinz, A.; Vogt, T.; Birkemeyer, C.; Wessjohann, L.; Frolov, A. Osmotic stress is accompanied by protein glycation in Arabidopsis thaliana. *J Exp Bot* **2016**, *67*, 6283–6295, doi:10.1093/jxb/erw395.

26.　Casey, R.; Domoney, C. Pea globulins. In *Seed proteins*; Shawrey, P.R., Casey, R., Eds.; Springer Netherlands: Dordrecht, 1999; pp. 171–208 ISBN 978-94-010-5904-6.

27.　Gaupels, F.; Furch, A.C.U.; Zimmermann, M.R.; Chen, F.; Kaever, V.; Buhtz, A.; Kehr, J.; Sarioglu, H.; Kogel, K.-H.; Durner, J. Corrigendum: Systemic Induction of NO-, Redox-, and cGMP Signaling in the Pumpkin

Extrafascicular Phloem upon Local Leaf Wounding. *Frontiers in Plant Science* **2016**, *7*, 281, doi:10.3389/fpls.2016.00281.

28. Min, C.W.; Gupta, R.; Kim, S.W.; Lee, S.E.; Kim, Y.C.; Bae, D.W.; Han, W.Y.; Lee, B.W.; Ko, J.M.; Agrawal, G.K.; Rakwal, R.; Kim, S.T. Comparative Biochemical and Proteomic Analyses of Soybean Seed Cultivars Differing in Protein and Oil Content. *J. Agric. Food Chem.* **2015**, *63*, 7134–7142, doi:10.1021/acs.jafc.5b03196.

29. Ogura, T.; Ogihara, J.; Sunairi, M.; Takeishi, H.; Aizawa, T.; Olivos-Trujillo, M.R.; Maureira-Butler, I.J.; Salvo-Garrido, H.E. Proteomic characterization of seeds from yellow lupin (Lupinus luteus L.). *Proteomics* **2014**, *14*, 1543–1546, doi:10.1002/pmic.201300511.

30. Krishnan, H.B.; Natarajan, S.S.; Oehrle, N.W.; Garrett, W.M.; Darwish, O. Proteomic Analysis of Pigeonpea (Cajanus cajan) Seeds Reveals the Accumulation of Numerous Stress-Related Proteins. *J. Agric. Food Chem.* **2017**, *65*, 4572–4581, doi:10.1021/acs.jafc.7b00998.

31. Miernyk, J.A.; Hajduch, M. Seed proteomics. *Journal of Proteomics* **2011**, *74*, 389–400, doi:10.1016/j.jprot.2010.12.004.

32. Yin, G.; Xin, X.; Fu, S.; An, M.; Wu, S.; Chen, X.; Zhang, J.; He, J.; Whelan, J.; Lu, X. Proteomic and Carbonylation Profile Analysis at the Critical Node of Seed Ageing in Oryza sativa. *Sci Rep* **2017**, *7*, doi:10.1038/srep40611.

33. Min, C.W.; Lee, S.H.; Cheon, Y.E.; Han, W.Y.; Ko, J.M.; Kang, H.W.; Kim, Y.C.; Agrawal, G.K.; Rakwal, R.; Gupta, R.; Kim, S.T. In-depth proteomic analysis of Glycine max seeds during controlled deterioration treatment reveals a shift in seed metabolism. *J Proteomics* **2017**, *169*, 125–135, doi:10.1016/j.jprot.2017.06.022.

34. Trudgian, D.C.; Fischer, R.; Guo, X.; Kessler, B.M.; Mirzaei, H. GOAT – A simple LC-MS/MS gradient optimization tool. *Proteomics* **2014**, *14*, 1467–1471, doi:10.1002/pmic.201300524.

35. Waas, M.; Bhattacharya, S.; Chuppa, S.; Wu, X.; Jensen, D.R.; Omasits, U.; Wollscheid, B.; Volkman, B.F.; Noon, K.R.; Gundry, R.L. Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins. *Anal Chem* **2014**, *86*, 1551–1559, doi:10.1021/ac403185a.

36. Wiśniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362, doi:10.1038/nmeth.1322.

37. Han, C.; Yin, X.; He, D.; Yang, P. Analysis of Proteome Profile in Germinating Soybean Seed, and Its Comparison with Rice Showing the Styles of Reserves Mobilization in Different Crops. *PLOS ONE* **2013**, *8*, e56947, doi:10.1371/journal.pone.0056947.

38. Fedorova, M.; Frolov, A.; Hoffmann, R. Fragmentation behavior of Amadori-peptides obtained by non-enzymatic glycosylation of lysine residues with ADP-ribose in tandem mass spectrometry. *J Mass Spectrom* **2010**, *45*, 664–669, doi:10.1002/jms.1758.

39. Matamoros, M.A.; Kim, A.; Peñuelas, M.; Ihling, C.; Griesser, E.; Hoffmann, R.; Fedorova, M.; Frolov, A.; Becana, M. Protein Carbonylation and Glycation in Legume Nodules. *Plant Physiology* **2018**, pp.00533.2018, doi:10.1104/pp.18.00533.

40. Gomes, L.S.; Senna, R.; Sandim, V.; Silva-Neto, M.A.C.; Perales, J.E.A.; Zingali, R.B.; Soares, M.R.; Fialho, E. Four conventional soybean [Glycine max (L.) Merrill] seeds exhibit different protein profiles as revealed by proteomic analysis. *J. Agric. Food Chem.* **2014**, *62*, 1283–1293, doi:10.1021/jf404351g.

41. Berjak, P.; Pammenter, N.W. From Avicennia to Zizania: seed recalcitrance in perspective. *Ann. Bot.* **2008**, *101*, 213–228, doi:10.1093/aob/mcm168.

42. Ellis, R.H. The encyclopaedia of seeds: science, technology and uses. *Ann Bot* **2007**, *100*, 1379, doi:10.1093/aob/mcm225.

43. Wang, X.; Tang, D.; Huang, D. Proteomic analysis of pakchoi leaves and roots under glycine–nitrogen conditions. *Plant Physiology and Biochemistry* **2014**, *75*, 96–104, doi:10.1016/j.plaphy.2013.12.012.

44. Smolikova, G.; Kreslavski, V.; Shiroglazova, O.; Bilova, T.; Sharova, E.; Frolov, A.; Medvedev, S. Photochemical activity changes accompanying the embryogenesis of pea (Pisum sativum) with yellow and green cotyledons. *Functional Plant Biol.* **2018**, *45*, 228–235, doi:10.1071/FP16379.

45. Smolikova, G.N.; Medvedev, S.S. Photosynthesis in the seeds of chloroembryophytes. *Russ J Plant Physiol* **2016**, *63*, 1–12, doi:10.1134/S1021443715060163.

46. Cuming, A.C. LEA Proteins. In *Seed Proteins*; Shewry, P.R., Casey, R., Eds.; Springer Netherlands: Dordrecht, 1999; pp. 753–780 ISBN 978-94-011-4431-5.

47.   Müntz, K.; Belozersky, M.A.; Dunaevsky, Y.E.; Schlereth, A.; Tiedemann, J. Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. *J Exp Bot* **2001**, *52*, 1741–1752, doi:10.1093/jexbot/52.362.1741.

48.   Hellmann, H.; Estelle, M. Plant Development: Regulation by Protein Degradation. *Science* **2002**, *297*, 793–797, doi:10.1126/science.1072831.

49.   Bilova, T.; Lukasheva, E.; Brauch, D.; Greifenhagen, U.; Paudel, G.; Tarakhovskaya, E.; Frolova, N.; Mittasch, J.; Balcke, G.U.; Tissier, A.; Osmolovskaya, N.; Vogt, T.; Wessjohann, L.A.; Birkemeyer, C.; Milkowski, C.; Frolov, A. A Snapshot of the Plant Glycated Proteome: Structural, Functional and Mechanistic Aspects. *Journal of Biological Chemistry* **2016**, *291*, doi:10.1074/jbc.M115.678581.

50.   Catusse, J.; Meinhard, J.; Job, C.; Strub, J.-M.; Fischer, U.; Pestsova, E.; Westhoff, P.; Van Dorsselaer, A.; Job, D. Proteomics reveals potential biomarkers of seed vigor in sugarbeet. *Proteomics* **2011**, *11*, 1569–1580, doi:10.1002/pmic.201000586.

51.   Wang, L.; Ma, H.; Song, L.; Shu, Y.; Gu, W. Comparative proteomics analysis reveals the mechanism of pre-harvest seed deterioration of soybean under high temperature and humidity stress. *J Proteomics* **2012**, *75*, 2109–2127, doi:10.1016/j.jprot.2012.01.007.

52.   Gallardo, K.; Job, C.; Groot, S.P.; Puype, M.; Demol, H.; Vandekerckhove, J.; Job, D. Proteomic analysis of arabidopsis seed germination and priming. *Plant Physiol.* **2001**, *126*, 835–848.

53.   Yacoubi, R.; Job, C.; Belghazi, M.; Chaibi, W.; Job, D. Toward characterizing seed vigor in alfalfa through proteomic analysis of germination and priming. *J. Proteome Res.* **2011**, *10*, 3891–3903, doi:10.1021/pr101274f.

54.   Rajjou, L.; Lovigny, Y.; Groot, S.P.C.; Belghazi, M.; Job, C.; Job, D. Proteome-wide characterization of seed aging in Arabidopsis: a comparison between artificial and natural aging protocols. *Plant Physiol.* **2008**, *148*, 620–641, doi:10.1104/pp.108.123141.

55.   Yamauchi, Y.; Ejiri, Y.; Toyoda, Y.; Tanaka, K. Identification and biochemical characterization of plant acylamino acid-releasing enzyme. *J. Biochem.* **2003**, *134*, 251–257.

56.   Gomes Marcelo; Garcia Queila Reactive oxygen species and seed germination. *biolog* **2013**, *68*, 351, doi:10.2478/s11756-013-0161-y.

57.   Sharma, P.; Jha, A.B.; Dubey, R.S.; Pessarakli, M. Reactive Oxygen Species, Oxidative Damage, and Antioxidative Defense Mechanism in Plants under Stressful Conditions. *Journal of Botany* **2012**, *2012*, e217037, doi:10.1155/2012/217037.

58.   Gallardo, K.; Le Signor, C.; Vandekerckhove, J.; Thompson, R.D.; Burstin, J. Proteomics of Medicago truncatula Seed Development Establishes the Time Frame of Diverse Metabolic Processes Related to Reserve Accumulation. *Plant Physiol* **2003**, *133*, 664–682, doi:10.1104/pp.103.025254.

59.   Wang, C.; Yue, W.; Ying, Y.; Wang, S.; Secco, D.; Liu, Y.; Whelan, J.; Tyerman, S.D.; Shou, H. Rice SPX-Major Facility Superfamily3, a Vacuolar Phosphate Efflux Transporter, Is Involved in Maintaining Phosphate Homeostasis in Rice. *Plant Physiol.* **2015**, *169*, 2822–2831, doi:10.1104/pp.15.01005.

60.   Yin, X.; He, D.; Gupta, R.; Yang, P. Physiological and proteomic analyses on artificially aged Brassica napus seed. *Front Plant Sci* **2015**, *6*, doi:10.3389/fpls.2015.00112.

61.   Xin, X.; Lin, X.-H.; Zhou, Y.-C.; Chen, X.-L.; Liu, X.; Lu, X.-X. Proteome analysis of maize seeds: the effect of artificial ageing. *Physiologia Plantarum* **2011**, *143*, 126–138, doi:10.1111/j.1399-3054.2011.01497.x.

62.   Stiti, N.; Missihoun, T.D.; Kotchoni, S.O.; Kirch, H.-H.; Bartels, D. Aldehyde Dehydrogenases in Arabidopsis thaliana: Biochemical Requirements, Metabolic Pathways, and Functional Analysis. *Front Plant Sci* **2011**, *2*, doi:10.3389/fpls.2011.00065.

63.   Sauter, M. Phytosulfokine peptide signalling. *J. Exp. Bot.* **2015**, *66*, 5161–5169, doi:10.1093/jxb/erv071.

64.   Matsuoka, K.; Demura, T.; Galis, I.; Horiguchi, T.; Sasaki, M.; Tashiro, G.; Fukuda, H. A Comprehensive Gene Expression Analysis Toward the Understanding of Growth and Differentiation of Tobacco BY-2 Cells. *Plant Cell Physiol* **2004**, *45*, 1280–1289, doi:10.1093/pcp/pch155.

65.   *ISTA. 2018. International rules for seed testing*; International Seed Testing Association: Basserdorf, Switzerland, 2018; ISBN ISSN 2310-3655.

66.   Matthews, S.; Powell, A. Electrical Conductivity Vigour Test: Physiological Basis and Use. *SeedTestInt* **2006**, 32–35.

67.   Lichtenthaler, H.K.; Wellburn, A.R. Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. *Biochemical Society Transactions* **1983**, *11*, 591–592, doi:10.1042/bst0110591.

68.   Bilova, T.; Paudel, G.; Shilyaev, N.; Schmidt, R.; Brauch, D.; Tarakhovskaya, E.; Milrud, S.; Smolikova, G.; Tissier, A.; Vogt, T.; Sinz, A.; Brandt, W.; Birkemeyer, C.; Wessjohann, L.A.; Frolov, A. Global proteomic

analysis of advanced glycation end products in the Arabidopsis proteome provides evidence for age-related glycation hot spots. *J. Biol. Chem.* **2017**, *292*, 15758–15776, doi:10.1074/jbc.M117.794537.

69. Greifenhagen, U.; Nguyen, V.D.; Moschner, J.; Giannis, A.; Frolov, A.; Hoffmann, R. Sensitive and site-specific identification of carboxymethylated and carboxyethylated peptides in tryptic digests of proteins and human plasma. *J. Proteome Res.* **2015**, *14*, 768–777, doi:10.1021/pr500799m.

70. Greifenhagen, U.; Frolov, A.; Blüher, M.; Hoffmann, R. Plasma Proteins Modified by Advanced Glycation End Products (AGEs) Reveal Site-specific Susceptibilities to Glycemic Control in Patients with Type 2 Diabetes. *J. Biol. Chem.* **2016**, *291*, 9610–9616, doi:10.1074/jbc.M115.702860.

71. Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **2010**, *26*, 680–682, doi:10.1093/bioinformatics/btq003.

72. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **2017**, *45*, D158–D169, doi:10.1093/nar/gkw1099.

73. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **1999**, *27*, 29–34.

74. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457-462, doi:10.1093/nar/gkv1070.

75. Usadel, B.; Poree, F.; Nagel, A.; Lohse, M.; Czedik-Eysenberg, A.; Stitt, M. A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant Cell Environ.* **2009**, *32*, 1211–1229, doi:10.1111/j.1365-3040.2009.01978.x.

76. King, B.R.; Vural, S.; Pandey, S.; Barteau, A.; Guda, C. ngLOC: software and web server for predicting protein subcellular localization in prokaryotes and eukaryotes. *BMC Res Notes* **2012**, *5*, 351, doi:10.1186/1756-0500-5-351.

77. Heberle, H.; Meirelles, G.V.; da Silva, F.R.; Telles, G.P.; Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **2015**, *16*, doi:10.1186/s12859-015-0611-3.