SEED PROTEOME ANALYSIS AND PROTEOME DYNAMICS DURING SEED GERMINATION

HABANOVA HANA
Department of Molecular Biology and Radiobiology
Mendel University in Brno
Zemedelska 1, 613 00 Brno
CZECH REPUBLIC

habanova.h@seznam.cz

Abstract: Despite the huge progress that has been made in the last decade, the molecular mechanisms regulating seed germination and early seed development are far from being resolved. Induction of metabolic genes involved in germination starts around 12 hours after imbibition. Thus, most of the early events are mediated by molecules stored in the seed during maturation and are not accessible to transcriptomic analyses. Proteome analysis has been extensively employed in the past but the coverage of observed seed proteome is relatively low even in present-day high-impact studies. Here, we analysed proteome of two model species, *Arabidopsis thaliana* and barley (*Hordeum vulgare*). We employed several complementary approaches to increase the proteome coverage and build a library suitable for targeted protein quantitation. The combination of fractionations and an alternative MS/MS data processing significantly improved our detection limits. Our results indicate that the seed proteome coverage is limited not only by extraction efficiency or depletion of abundant proteins, but also by an inadequate spectral data interpretation.

Key Words: seed proteomics, mass spectrometry, proteome fractionation

INTRODUCTION

Germination is a crucial process which affects viability and productivity of plants. In terms of physiology, germination is quite well described. However, in terms of molecular biology it still remains unclear. Germination is defined as a three-phase process which begins with intensive water intake and which is ended by testa and endosperm rupture. The first phase of germination of non dormant seeds is characterized by metabolism activation after achieving approximately 60% of hydration. Induction of metabolic genes involved in germination starts around 8 hours after imbibition (Rajjou et al. 2012). Thus, most of the early events are mediated by molecules stored in the seed during maturation and are not accessible to transcriptomic analyses. Therefore, proteomic analysis should be the method of choice to understand molecular mechanism regulating seed germination and early seed development.

Proteome analysis has been extensively employed in the past but the coverage of observed seed proteome is relatively low even in present-day high-impact studies. Moreover, genome of majority of agriculturally important crops has not yet been fully sequenced and thus the databases for proteome annotation contain only small sets of well validated proteins. The techniques to increase the proteome coverage are available, but usually not optimized. Here, we tested two different approaches to reach a higher proteome coverage: (i) proteome fractionation and (ii) in silico reprocessing of HRMS (high resolution mass spectrometry) data.

MATERIAL AND METHODS

Plant material, cultivation and total protein extraction

Seeds of *Arabidopsis thaliana* (Col-0) and *Hordeum vulgare* (variety Sebastian) were imbibed with distilled water and harvested after 0-24 h of imbibition, frozen in liquid nitrogen and homogenized using a Retsch Mill MM400. The total protein was extracted by acetone/TCA/phenol extraction as described previously (Černý et al. 2013). In brief, homogenized tissue was extracted overnight with 10% (w/v) TCA in acetone (2 ml, -20°C), washed with 10% (w/v) TCA in distilled water then 80% (v/v) acetone, resuspended in 0.8 ml SDS buffer [2% (w/v) SDS, 30% (w/v) sucrose, 5% (v/v)

β-mecraptoethanol, 5 mM EDTA, 100 mM Tris, pH 8.0], and protein was extracted by 0.4 ml buffer-saturated phenol. Phenolic phase was collected and protein was precipitated overnight in 1.6 ml ice-cold 100 mM ammonium acetate in methanol (-20°C). Protein pellets were washed with 1.0 ml 80% (v/v) acetone in distilled water, dried and stored at -80°C until used.

Off-Gel Fractionation

Barley total protein extracts prepared as described above were dissolved in OFFGEL Stock Solution (thiourea 2 mM, DTT 60 mM, 10% (v/v) glycerol, ampholytes pH 3-10), loaded into wells with 24 cm IPG strips (pH 3-10, nonlinear) and processed according to the manufacturer's instructions (Offgel Fractionator 3100, Agilent). The resulting fractions were collected and then digested in solution with trypsin.

LC-MS analysis

Arabidopsis samples were prepared as described previously (Baldrianová et al. 2015). In brief, dried protein pellets were dissolved in 100 mM NH_4HCO_3, 8 M urea (400 µl). The protein concentration was estimated by the Bradford assay (Sigma-Aldrich), samples were diluted with acetonitrile in 100 mM NH_4HCO_3 to the final concentration 5% acetonitrile, 2M urea, 50 mM NH4HCO3 and subjected to in-solution digestion with immobilized trypsin beads (Promega; 3 µl beads per 100 µg of protein) at 37°C overnight. The resulting peptides were desalted (SPEC plate C18, Agilent), dried and dissolved in 0.5% (v/v) formic acid in 5% (v/v) acetonitrile, then analysed online by nanoflow C18 reverse-phase liquid chromatography using a 15 cm Ascentis Express Column (0.1 mm inner diameter; Sigma-Aldrich) and a Dionex Ultimate 3000 RSLC nano UPLC system (Thermo) directly coupled to a nanoESI source CaptiveSpray and an UHR maXis impact q-TOF mass spectrometer (Bruker). Peptides were eluted with a 120-min, 4% to 35% acetonitrile gradient (Novák et al. 2015).

Data processing

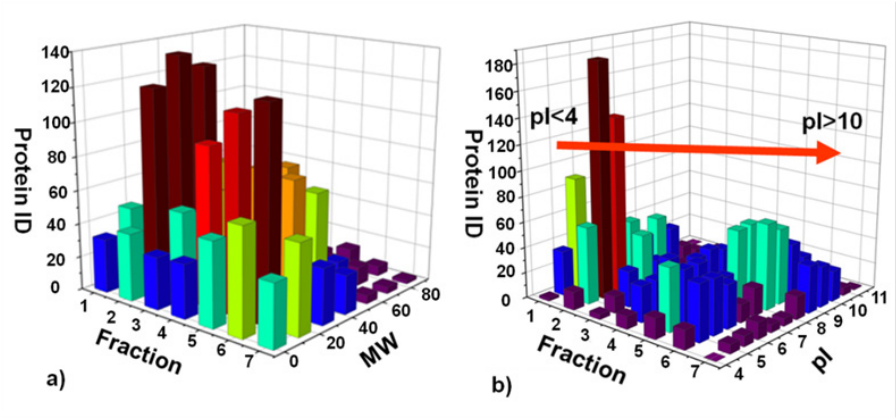
Peptide spectra were preprocessed with DataAnalysis (Bruker) and searched against barley and *Arabidopsis* TAIR10 protein databases using the Mascot algorithm and Bruker's ProteinScape inbuilt percolator algorithm (target FDR<1%). Skyline, Search GUI (1.30.1) and Peptide Shaker (0.41) (Vaudel et al. 2011) were used for a further spectra analyses and processing.

RESULTS AND DISCUSSION

Off-Gel separation of barley grain proteins

There are several suitable methods of proteome fractionation. Seeds are formed to provide nutrition for the embryo, thus the storage compounds (including storage proteins) represent the greatest portion of a seed's mass. The abundant storage proteins complicate analysis and interfere with the detection of lower abundant proteins. The fractionation on tissue-level (e.g. microdissection) is possible, but very demanding. Further, for some species, including *Arabidopsis*, the methodology does not allow rapid harvest of a sufficient amount of material for protein extraction. A more accessible is the fractionation on the protein level. The most common is the use of liquid chromatography or protein electrophoresis. The Off-gel fractionation method is an electrophoretic method based on isoelectric focusing and enables separation of proteins in a solution. Here, 1 mg of barley seed protein was fractionated into 12 fractions, but only seven fractions had a sufficient amount of protein for a further analysis (>100 µg). Fractions were analysed via LC-MS and MS/MS spectra were searched against barley database. The analysis of the theoretical molecular weight (MW) showed that the majority of proteins in all fractions have MW between 10 to 50 kDa (Figure 1a). The distribution of theoretical isoelectric points illustrates the efficiency of the separation and indicates a presence of proteoforms and/or post-translational modifications (Figure 1b). In accordance, there was a significant overlap between proteins identified in individual fractions. In total, 951 (a high confidence gene model version) and 561 (a low confidence gene models) were identified. This represented some 30% increase compared to the standard shot-gun approach. Though the peptide-based fractionation methods like strong cation exchange chromatography (SCX) would have better fractionation efficiency, the Off-gel separation retains the information about different proteoforms in the sample.

Figure 1 Distribution of proteins according to molecular weight (a) and isoelectric point (b) after Off-Gel separation



Advanced processing of mass spectrometry data

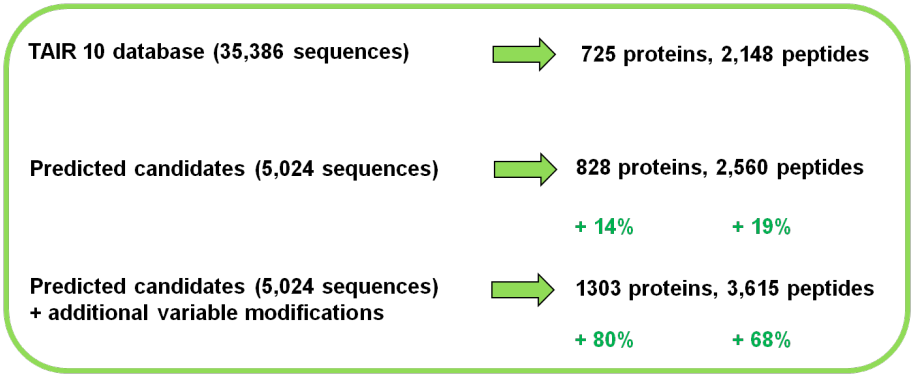
Proteomic analysis of barley seeds rely on still poorly annotated genome. Therefore barley is not the perfect candidate for digging in MS/MS spectra, increasing variable modifications and pin-pointing new peptide spectral matches. However, in the case of model plant *Arabidopsis thaliana*, we can use the available bioinformatics to increase the number of identifications in complex sample. *A. thaliana* seeds were prepared as is described in Material and Methods. Classical shotgun analysis followed by a gold-standard Mascot search engine resulted in identification of 1,450 proteins (the summary from several analyses). However, the Skyline analysis indicated that the 4,899 peptides used for protein identification by Mascot can be assigned to more than 7,000 known proteoforms in *Arabidopsis* proteome. These proteoforms correspond to 5,752 unique genes and 5,024 of them contain unique proteotypic peptides (Table 1). In theory, all these proteins could be present in the sample and they should be excluded only if there is no evidence of their proteotypic peptides.

Table 1 MS data post-processing

Number of detected proteins (peptides)	Search Algorithm
1.450 (4.899)	MASCOT
7.038	proteoforms
5.752	unique genes
5.024	proteins with proteotypic peptides

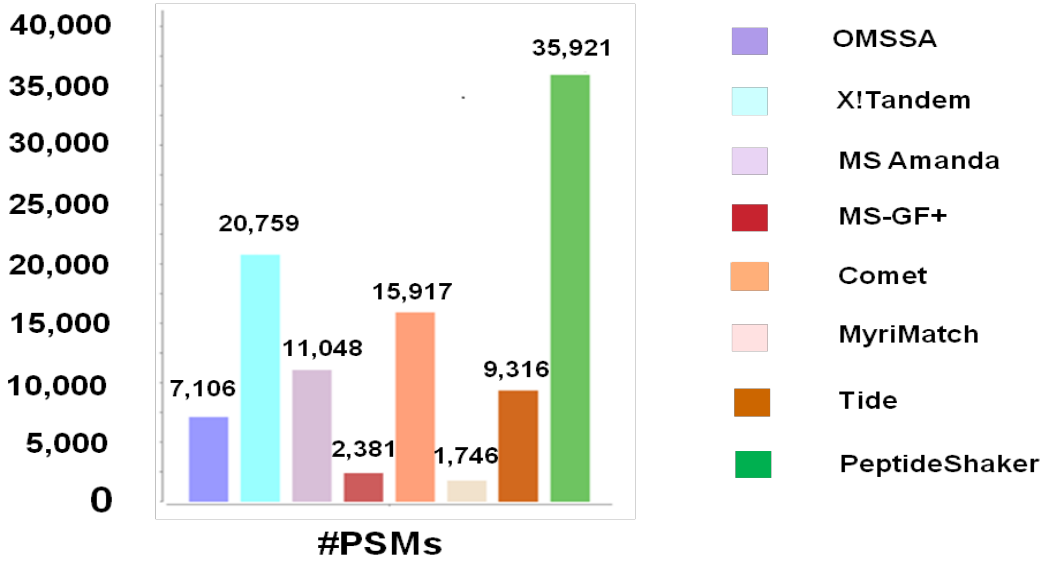
To test this theory, we used the sequences of these proteins as the Mascot database. As the number of detectable proteins is limited by the size of the reference database, by decreasing its size by more than 80% we were able to detect 828 proteins in a single analysis (14% increase). Moreover, we increased this number even further when we included additional variable modifications into the search parameters (1303 proteins identified) (Figure 2).

Figure 2 Backward analysis with database consisting of predicted candidate proteins as a useful approach to increase number of detectable proteins



As the next step, we complemented Mascot results with that of seven alternative search engines and the resulting data were combined in PeptideShaker. In total, we reached almost 36,000 peptide spectral matches (PSMs) (Figure 3).

Figure 3 Combination of data resulting from different search algorithms provides increasing number of peptide spectral matches (PSMs)



CONCLUSION

Seed proteomic analysis is a promising tool to study the molecular mechanism regulating seed germination and early seed development. However, seed proteome analysis is still difficult and limited by many obstacles. Here, we show the benefits of fractionation and bioinformatics in analysis of barley and *Arabidopsis* seed proteome. Our data will serve as the protein library and will be used for a targeted proteomic analysis of seed germination.

ACKNOWLEDGEMENT

This work was supported by grants P305/12/2144 (CSF), TE02000177 (TACR), funds from the ERDF for 'CEITEC–Central European Institute of Technology' (CZ.1.05/1.1.00/02.0068).

REFERENCES

Baldrianová J., Černý M., Novák J., Jedelský P. L., Divíšková E., Brzobohatý B. 2015. Arabidopsis proteome responses to the smoke-derived growth regulator karrikin. *Journal of Proteomics*, 120: 7–20.

Černý M., Kuklová A., Hoehenwarter W., Fragner L., Novák O., Rotková G., Jedelský P. L., Žáková K., Šmehilová M., Strnad M., Weckwerth W., Brzobohatý B. 2013a. Proteome and metabolome profiling of cytokinin action in Arabidopsis identifying both distinct and similar responses to cytokinin down- and up-regulation. *Journal of Experimental Botany,* 64(14): 4193–4206.

Novák J., Černý M., Pavlů J., Zemánková J., Skalák J., Plačková L., Brzobohatý B. 2015. Roles of proteome dynamics and cytokinin signaling in root-to-hypocotyl ratio changes induced by shading roots of Arabidopsis seedlings. *Plant Cell Physiology,* 56(5): 1006–1018.

Rajjou L., Gallardo K., Catusse J., Bally J., Job C., Job D. 2012. Seed Germination and Vigor. *The Annual Review of Plant Biology,* 63: 507–533.

Vaudel M., Barsnes H., Berven F. S., Sickmann A., Martens L. 2011. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X! Tandem searches. *Proteomics*, 10(5): 996–999.