

UNcleProt (Universal Nuclear Protein database of barley): The first nuclear protein database that distinguishes proteins from different phases of the cell cycle

Nicolas Blavet ^a, Jana Uřinová^b, Hana Jeřábková^a, Ivo Chamrád^b, Jan Vrána ^a, René Lenobel^b, Jana Beinhauer^b, Marek Šebela^b, Jaroslav Doležel ^a, and Beáta Petrovská ^a

^aInstitute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic;

^bDepartment of Protein Biochemistry and Proteomics, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic

ABSTRACT

Proteins are the most abundant component of the cell nucleus, where they perform a plethora of functions, including the assembly of long DNA molecules into condensed chromatin, DNA replication and repair, regulation of gene expression, synthesis of RNA molecules and their modification. Proteins are important components of nuclear bodies and are involved in the maintenance of the nuclear architecture, transport across the nuclear envelope and cell division. Given their importance, the current poor knowledge of plant nuclear proteins and their dynamics during the cell's life and division is striking. Several factors hamper the analysis of the plant nuclear proteome, but the most critical seems to be the contamination of nuclei by cytosolic material during their isolation. With the availability of an efficient protocol for the purification of plant nuclei, based on flow cytometric sorting, contamination by cytoplasmic remnants can be minimized. Moreover, flow cytometry allows the separation of nuclei in different stages of the cell cycle (G1, S, and G2). This strategy has led to the identification of large number of nuclear proteins from barley (*Hordeum vulgare*), thus triggering the creation of a dedicated database called UNcleProt, <http://barley.gambrinus.ueb.cas.cz/>.

ARTICLE HISTORY

Received 29 June 2016

Revised 20 October 2016

Accepted 25 October 2016

KEYWORDS

barley; cell cycle; database; flow-cytometry; localization; mass spectrometry; nuclear proteome; nucleus

Introduction

The nucleus is a key feature of eukaryotic cells and is characterized by the high complexity and dynamic organization of its components. Inside the nucleus, the nuclear genome is organized both in a hierarchical manner and within discrete territories. In addition to DNA, the cell nucleus contains nuclear bodies and specialized sub-nuclear compartments (for a review see refs. ^{1,2}). Nuclear compartmentalization reflects the spatial arrangement of the genome and the DNA-related processes that occur in this organelle. Next to DNA and RNA, the most abundant class of molecules present in the nucleus are nuclear proteins.³ They play a major role in DNA assembly and its packing in the small space of the nuclear volume. The complex network of nuclear proteins performs diverse functions that are essential for maintaining dynamically

changing genome organization and regulating of gene expression. Proteins also form the main building blocks of nuclear membrane pores,⁴ constitute lamina or lamina-like structures that modulate the size and shape of the nucleus,⁵ play an important role in chromatin organization and gene expression, and connect the nuclear lamina to the cytoskeleton, which is necessary for both nuclear positioning and migration.⁶

Though proteins represent an abundant and indispensable part of the nucleus, knowledge of the plant nuclear proteome remains limited. For that reason, scientists associated with the International Plant Nucleus Consortium⁷ have, since its establishment, been focusing on the detailed characterization of specific nuclear proteins and overall nuclear protein composition to obtain more information about the complex nuclear machinery.

CONTACT Beáta Petrovská  petrovska@ueb.cas.cz  Institute of Experimental Botany AS CR, Centre of Plant Structural and Functional Genomics, Slechtitelu 31, CZ-783 71 Olomouc, Czech Republic.

© 2017 Nicolas Blavet, Jana Uřinová, Hana Jeřábková, Ivo Chamrád, Jan Vrána, René Lenobel, Jana Beinhauer, Marek Šebela, Jaroslav Doležel, and Beáta Petrovská. Published with license by Taylor & Francis.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

The first list of identified plant nuclear proteins was published in 2003 in connection with the sequenced genome of *Arabidopsis thaliana*.^{8,9} Since then, the nuclear proteomes from different types of tissues have been studied in several species (for a review, see ref.²), including *Arabidopsis*,¹⁰⁻¹² rice,¹³⁻¹⁹ hot pepper,²⁰ chickpea,²¹⁻²⁴ barrel clover,²⁵ maize,²⁶⁻²⁷ black-stick lily,^{28,29} soybean³⁰ and, just recently, wheat.^{31,32} The main aim of these studies was the analysis of nuclear proteome changes in response to abiotic and biotic stresses.

Petrovská et al.³³ used a new approach to purify the cell nuclei of barley (*Hordeum vulgare*) and identified 803 proteins in G1-phase nuclei. In a subsequent study, the team analyzed proteins from S- and G2-phase nuclei, which led to a dramatic increase in the total number of identified nuclear proteins of barley (Uřinová et al., unpublished data). To organize this large amount of data, a database system of barley nuclear proteins became necessary.

Here, we report on the creation of a barley (*H. vulgare* L., cv. Morex) nuclear protein database (UNcleProt, <http://barley.gambrinus.ueb.cas.cz/>) that provides information about the barley nuclear proteins that have been identified to date. The acronym UNcleProt is based on a truncated anagram “uncle” from the term “nuclear.” Besides containing the anagram, the term UNcleProt has a self-explanatory feature: U stands for “Universal” (3 phases of the cell cycle are covered), “Ncle” for “Nuclear” and “Prot” for proteins. UNcleProt is available freely. In addition to survey sequences,³⁴ a reference barley genome with fully assembled and annotated chromosomes is being completed, and the barley nuclear proteome database will be its perfect complement. Together, the 2 resources will represent one of the most complete data source on the barley nuclear proteome.

Results

The UNcleProt database includes 6 main pages: Home, Browse, Query, Blast, Download and Remarks, which are accessible via the navigation bar. The Home page provides information on the database with a short summary and 2 pictures: a Venn diagram showing the number of proteins identified thus far in each phase of the cell cycle (Fig. 1), the number of peptides detected, and a graph showing the most abundant gene ontology (GO) terms, available with Uniprot

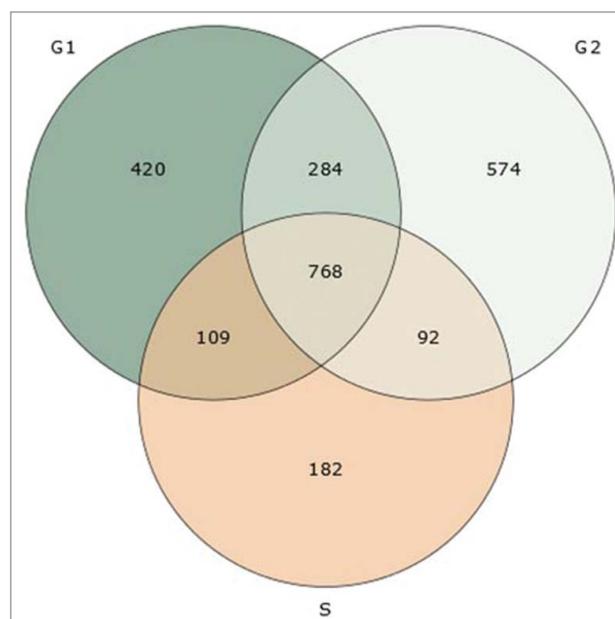


Figure 1. A Venn diagram showing the number of barley nuclear proteins identified in 3 different phases of cell cycle (unique and shared identifications).

protein annotations, for each cell cycle phase (Fig. 2). A link to the first related article from our laboratory by Petrovská et al.³³ is also provided. The Browse page contains a table with all proteins identified in the nuclei from different phases of the cell cycle (G1, S and G2) and shows their accession numbers and names, provides information on the cell cycle phase during which the proteins occur and finally presents a link to access the respective Protein and Peptide Information page for each selected protein (described below). The Query page (Fig. 3A) allows the user to search the database for accession numbers or keywords, and the result is displayed as a table similar to the one generated by the Browse page (Fig. 3B). The Blast page makes it possible to run a BLAST search³⁵ using the blastp or blastx programs against each identified protein sequence. The Download page allows the set of proteins identified for each cell cycle phase to be retrieved either separately or all together. Finally, the Remarks page provides technical information on the mass spectrometry results.

In addition to the primary page layout, the Protein and Peptide Information page provides detailed information on the identified proteins (Fig. 4). The data are presented in 2 tables: the first summarizes information available from the UniProt database, such as the ID number with a hyperlink, protein name, sequence length, keywords, database

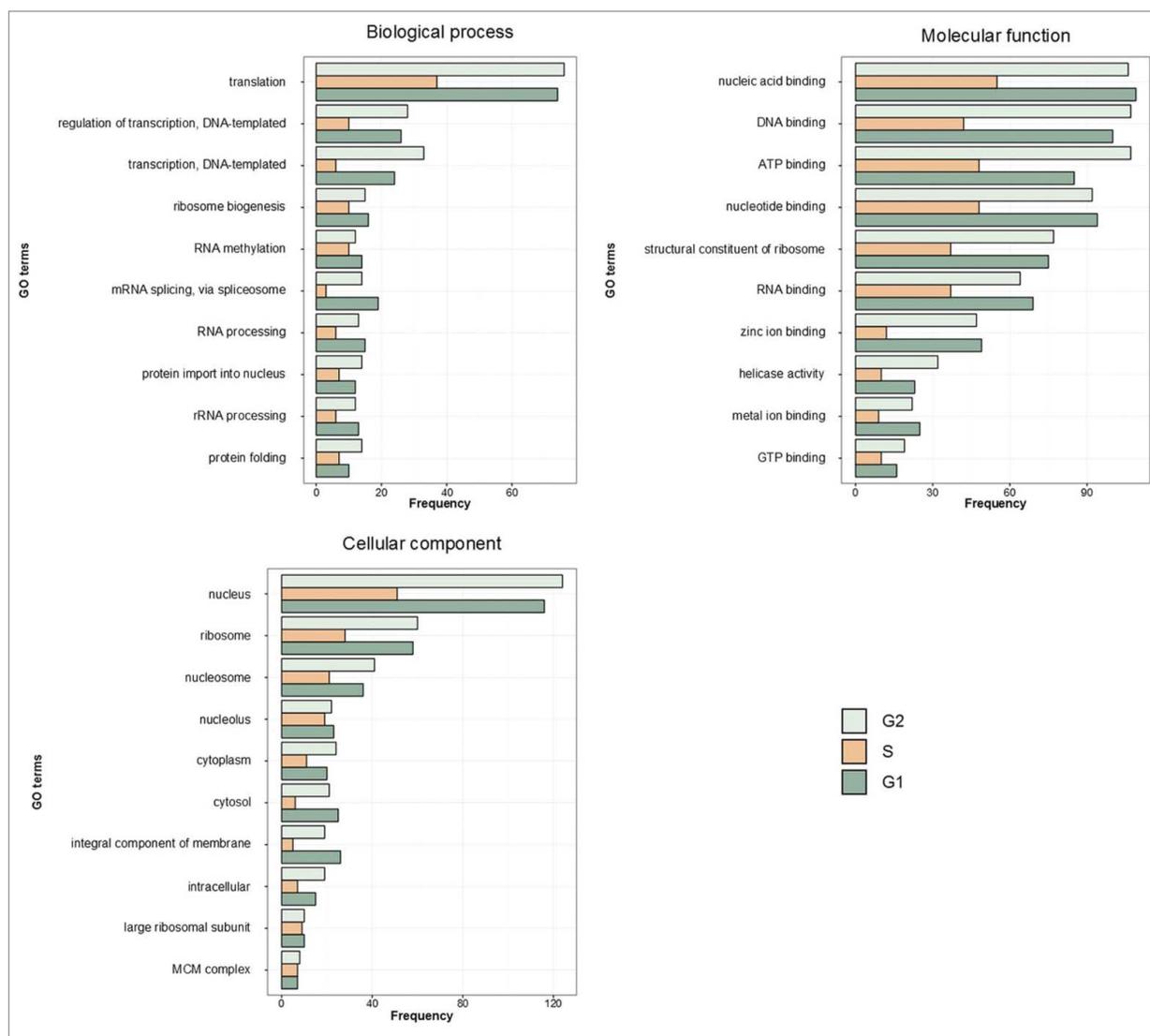


Figure 2. Distribution of the 10 most abundant GO terms in all 3 categories: biological process, molecular function and cellular component. GO terms were extracted from Uniprot annotations. The results are shown separately for each of the cell cycle phases.

cross-references such as GO and protein domain IDs, and then a sequence coverage value based on the identification by mass spectrometry (MS) and the whole amino acid sequence, in which the coverage by sequenced peptides is highlighted in red. Information on the availability of a known 3D structure is accessible through links to the appropriate databases: the Database of Comparative Protein Structure Models (ModBase³⁶) and the Database of Protein Disorder and Mobility Annotations (MobiDB, created by Bio-computingUP Lab, University of Padua, Italy). However, because a majority of the barley amino acid sequences have not yet been used for structure modeling, a model calculation in the ModBase usually takes approximately 2 d.

Most of the identified barley nuclear proteins are annotated as uncharacterized or predicted [62% (1505) and 24% (594), respectively] in the UniProt. For such proteins, a table row called GenBank might contain the 10 best hits obtained from the GenBank nucleotide database (nt). At the end of the table, a summary of the MS results with the corresponding Mascot score and the number of peptides detected for each cell cycle phase is presented. Importantly, a vast majority of uncharacterized or predicted proteins could be annotated by looking for orthologous sequences using blastp³⁵ or conserved domain³⁷ searches and this information will gradually be added and will appear in new releases of the database. Only less than 5 % of the total number of proteins identified

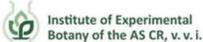
A

UNcieProt

Home Browse Query Blast Download Remarks

Enter accessions or keywords.
 accession e.g. F2CQ27,F2CQ57 F2CQF3
 keyword e.g. histone, kinet*, GO:0005634

Search for





This website uses cookies

B

UNcieProt

Home Browse Query Blast Download Remarks

Link	Accession	Protein Name	G1	S	G2
<input type="button" value="info"/>	Q9M3X1	Putative kinetochore protein OS=Hordeum vulgare GN=skp1 PE=2 SV=1	1	0	0
<input type="button" value="info"/>	Q9M3X0	Putative kinetochore protein OS=Hordeum vulgare GN=cbf5 PE=2 SV=1	0	1	0
<input type="button" value="info"/>	M0X4Q2	Uncharacterized protein OS=Hordeum vulgare var. distichum PE=4 SV=1	0	0	1





This website uses cookies

Figure 3. Query search and result pages: (A) a screenshot of the Query page, where searching is possible by either accession number or keyword; (B) a screenshot of the Query page with a table containing database search results.

in each cell cycle phase have not yet been provided with at least such an annotation.

To validate the subcellular localization of the unknown, predicted or uncharacterized proteins identified by MS, 2 subcellular and one subnuclear localization programs have been used. These predictions are shown in the first table (“Protein Information”). Based on the accuracy of 6 subcellular localization programs tested by Xiong et al.,³⁸ we chose Plant-mPLOC (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>)³⁹ and CELLO (<http://cello.life.nctu.edu.tw/>)⁴⁰. Plant-mPLOC employs sequence-based predictions (amino acid composition, functional domains, sequential evolution features, GO terms). CELLO relies on feature search methods employing amino acid composition, dipeptide composition,

portioned amino acid composition and the physicochemical properties of amino acids. The prediction accuracy for all identified barley nuclear proteins was monitored using the subnuclear web server Nuc-PLOC (<http://www.csbio.sjtu.edu.cn/bioinf/Nuc-PLOC/>)⁴¹. This tool specifically predicts nuclear localization with an accuracy of nearly 100 %.³⁸

The second table (“Peptide Information”) contains data on all identified peptides related to a given protein, such as the cell cycle phase of the source nuclei, mass spectrometer used, sequence, measured mass/charge ratio, measured mass and charge, deviation calculated for the experimental and theoretical mass, retention time (min) in the liquid chromatographic separation of the respective digest, precursor ion intensity, peptide score given by Mascot search engine, position in the



Home | Browse | Query | Blat | Download | Remarks

F2CQF3

Protein Information	
Uniprot ID	F2CQF3_HORV0 link to Uniprot
Protein name	Predicted protein OS=Hordeum vulgare var. distichum PE-2 SV=1
Length (aa)	475
Status	Unreviewed
Info	SubName: Full+Predicted protein [ECO:0000313EMBL_BAI85074.1]; SubName: Full+Uncharacterized protein [ECO:0000313EnsemblPlants_MLOC_58279.1];
Species	Hordeum vulgare var. distichum (Domesticated barley)
Database Cross-reference	EMBL: AK35385; BAI85074.1; -; mRNA EnsemblPlants: MLOC_58279.1; MLOC_58279.1; MLOC_58279.1; MLOC_58279.2; MLOC_58279.2; MLOC_58279.2; MLOC_58279.2 Proteomes: UP000011116; Unassembled WGS sequence GO: GO:0005524; F; ATP binding; EA:UaProPK-KW GO: GO:0003676; F; nucleic acid binding; EA:UaProPK-KW Gene3D: 3.40.50.300; -2; InterPro: IPR011545; DEAD/DEAH_box_helicase_dom. InterPro: IPR014801; Helicase_ATP_bd. InterPro: IPR001650; Helicase_C. InterPro: IPR021417; P-loop_NTPase. InterPro: IPR000629; RNA_helicase_DEAD_box_CS. InterPro: IPR014014; RNA_helicase_DEAD_Q_motif. Pfam: PF00270; DEAD_1. Pfam: PF00271; Helicase_C_1. SMART: SM00487; DEAD_1. SUPFAM: SSF5240; SSF5240_2. PROSITE: PS06809; DEAD_ATP_HELICASE_1. PROSITE: PS05192; HELICASE_ATP_BIND_1_1. PROSITE: PS05194; HELICASE_CTER_1. PROSITE: PS05195; Q_MOTIF_1.
Evidence	2 Evidence at transcript level
Subcellular/subnuclear localization prediction	CELLO: Mitochondrial 1 884% CELLO full results Plant-mPloc: Nucleus Nuc-Ploc: Nucleolus
Keywords	ATP-binding [ECO:0000256RUEBase_RU000492]; Complete proteome [ECO:0000313Proteomes_UP000011116]; Helicase [ECO:0000256RUEBase_RU000492]; Hydrolase [ECO:0000256RUEBase_RU000492]; Nucleotide-binding [ECO:0000256RUEBase_RU000492]; Reference proteome [ECO:0000313Proteomes_UP000011116]
Genbank	AK353855 Hordeum vulgare subsp. vulgare mRNA for predicted protein, complete cds, clone: NASH000221 AK333087 Triticum aestivum cDNA, clone: WT05_020, cultivar: Chinese Spring XR_731016 PREDICTED: Brachypodium distachyon DEAD-box ATP-dependent RNA helicase 30 (LOC100832027), transcript variant X2, misc_RNA XM_003564844 PREDICTED: Brachypodium distachyon DEAD-box ATP-dependent RNA helicase 30 (LOC100832027), transcript variant X1, mRNA XM_002459919 Sorghum bicolor hypothetical protein, mRNA XM_006486529 PREDICTED: Oryza brachyantha DEAD-box ATP-dependent RNA helicase 30-like (LOC102703183), mRNA XM_004979954 PREDICTED: Setaria italica DEAD-box ATP-dependent RNA helicase 30 (LOC101731311), mRNA NM_001185776 Oryza sativa Japonica Group Os11g0911100 (Os11g0911100), partial cds AK064782 Oryza sativa Japonica Group cDNA clone, J013000912, full insert sequence XR_000048 PREDICTED: Phoenix dactyloides DEAD-box ATP-dependent RNA helicase 20-like (LOC10319819), transcript variant X2, misc_RNA
3D structure	ModBase search ModRIP search
Sequence coverage	124 / 475 = 26.11%
Sequence	MNFFVADNRVADPSSVDRRSDLAAPILAPFVVAQNPYAPPTFMAAPVGGVGGGGG GYGGGMYGGRGGGGGGGGGFRGGGGGGGSGRGLDLSLFLKDFRGLPFKESFYV ECFAVMQAMSETEVAQVRLRDIIVGGRVYKPIRFFHEANFFDCMGAIAKSGFVEPTPI QAQQRPMALRGKDVIGIAGTSGSKTLYSLPLGLVYVGAQPRLEGGQDFVLLAFTRFLA VGIQAEATKTSSTSRITAFCTFYGAQKQPIRLDLRQVEIVLAFPSRLIDMLRAGHTLRA RVYVLLVLEADRLDMGFEPQIRKLAQREDRQLYSAITWREVEITLAFQFNPKYK MIGTAEIAKHSIQIQLVEIVSDREKYPRLSLDLMDGSRILIFFTKKCKVTRQLR MDGWFALSIHGDKAQSERDVLAEFKNGKSFMAATDVAARGLGMVTCINIRIML
Overall protein score (Masscut)	G1: 272.20 S: 338.60 Q2: 142.40
Number of peptides	G1: 4 S: 8 Q2: 3 Total: 15

Peptide Information												
Peptide	Sequence	MS/MS	Score	Rank	Protein							
S MALDI	Y..ILPGLVHYGAQPR..L	1356 7938	1356 7938	1	-0.0172	-12.69	35	246.4	107.60	209-221	1	
G1 MALDI	R..DGLDELLEKDFDPR..G	1559 7948	1559 7948	1	-0.0116	-7.45	50	687	71.60	96-109	0	
G1 MALDI	R..LEQDGPVLLAIPFR..E	1691 9687	1691 9687	1	-0.0003	-0.21	42.5	1807.3	69.40	222-237	0	
G1 MALDI	Y..ILPGLVHYGAQPR..L	1356 7926	1356 7926	1	-0.0184	-13.56	35.85	255.2	67.40	209-221	1	
S MALDI	Y..ILPGLVHYGAQPR..L	1356 8037	1356 8037	1	-0.0074	-5.42	43.8	2526.3	65.60	209-221	1	
G1 MALDI	K..TLVYILPOLVHYGAQPR..L	1821 038	1821 038	1	-0.0001	-0.05	40.5	5800.4	63.80	209-221	0	
G2 MALDI	Y..ILPGLVHYGAQPR..L	1356 8016	1356 8016	1	-0.0094	-6.91	34.25	15996.0	60.70	209-221	0	
G2 MALDI	K..ANHSIQIQLVEIVSDREKYP..L	2363 192	2363 192	1	-0.0182	-7.71	47.5	16654.6	59.70	369-388	1	
S MALDI	R..MLMGPEPQIR..K	1336 6441	1336 6441	1	0.0053	4	40.82	227.8	42.10	313-323	2	
S MALDI	R..DVIQIAET..G	817 4357	817 4357	1	0.0056	6.8	33.13	1710	32.40	193-200	0	
S MALDI	L..SLPEKDFR..G	959 5337	959 5337	1	0.0028	2.96	37.93	696.4	25.70	102-109	0	
S MALDI	R..YADPSYR..D	958 4203	958 4203	1	-0.0062	-6.43	27.8	841.6	24.00	9-16	0	
S MALDI	K..ILAQERFOR..Q	1081 6451	1081 6451	1	-0.0026	-2.38	33.67	398.3	22.80	325-333	0	
G2 MALDI	R..QVEIVLAFPR..L	1111 6397	1111 6397	1	-0.0072	-6.52	34.25	13980	22.00	277-287	0	
S MALDI	K..NKSFIMAAAT..D	1005 5024	1005 5024	1	-0.001	-0.96	27.53	870.8	17.80	447-456	1	Oxidation: 7







This website uses cookies

Figure 4. Protein and Peptide Information page. This page presents information related to a particular protein, including the corresponding MS data and amino acid sequences of the identified peptides.

amino acid sequence, number of missed cleavage (MCV) sites and chemical modifications, i.e., if there is a modified amino acid present compared to a reference sequence.

Discussion

The UNcleProt database described here is built on the protein identification results that we have achieved since introducing the flow cytometric sorting method for barley nuclear proteomics.³³ This approach results in low contamination by cytosolic compounds, allows the purification of nuclei from different phases of the cell cycle and is applicable even to mitotic chromosomes.^{42,43} Barley is a suitable model for studying the nuclear proteomes of crops because it is a diploid self-pollinating species with a range of genetic and genomic resources available. Moreover, protocols for genetic transformation⁴⁴ and a range of both mutant lines and TILLING populations⁴⁵ are available to facilitate functional analyses. Barley's large genome (1C = 5,428 Mbp⁴⁶) is reflected by its large nuclei and the high yield of proteins obtained from flow-sorted nuclei. Numerous and notable results have recently been obtained, including a draft genome sequence of barley.^{47,48} Finally, the reference genome sequence to be published in 2016 (N. Stein, personal communication) will provide a perfect match to the barley nuclear protein database.

This barley nuclear proteome database (UNcleProt) contains 2,429 proteins identified via both MALDI-MS (matrix assisted laser desorption ionization – mass spectrometry) and ESI-MS (electrospray ionization – mass spectrometry) from G1, S and G2 cell cycle phase nuclei (Fig. 1) (Uřinová et al., unpublished results). To complete this protein set, a total of 34,675 peptides have been identified and assigned to the corresponding protein sequences. To date, none of the existing nuclear databases such as the yeast nuclear database,⁴⁹ vertebrate nuclear database (<http://npd.hgu.mrc.ac.uk/user/>^{50,51}), rice nuclear database (<http://gene64.dna.affrc.go.jp/RPD/main.html>¹⁶), or TAIR (www.arabidopsis.org) incorporate the deposition of nuclear proteins identified in different stages of the cell cycle. Both protein level or localization changes and the related regulatory aspects across the cell cycle are of a big interest due to their connection with fundamental processes of cell biology (such as cell division and signal transduction) and the pathology of diseases.

To our best knowledge, cell cycle phase-related alterations in plant nuclear proteomes have not yet been explored in detail. For human cells, differential proteomics studies covering the cell cycle dependence of nuclear proteins are typically targeted on a specific subgroup or process.^{52,53} The dynamic proteomics approach involving fluorescent tagging and microscopy demonstrated cell-cycle dependence in concentration levels for 8 of the analyzed 124 nuclear proteins from a human lung cancer cell library.⁵⁴ In a large scale comparative experiment with G1-, G2- and S-phase nuclei, which provided the largest portion of protein identifications archived in the UNcleProt database, 266 nuclear proteins in total were found enriched at least 2-fold in different cell cycle phases based on a semiquantitative spectral counting approach. For this group of proteins, Fig. 5 summarizes GO terms describing associated biological processes and shows a distribution of the observed cell cycle phase-dependent changes. Many of the available annotations for the identified barley nuclear proteins lack any functional information in current versions of protein sequence databases. Of the whole set, 1/4 are only predicted proteins, and almost 2/3 are still uncharacterized. Nevertheless, numerous uncharacterized/predicted proteins could be annotated by blastp search, conserved domain search or associated with GO terms to tentatively describe the molecular function, biological process and cellular component. Not surprisingly, most of the barley nuclear proteins have DNA- or nucleic acid-binding functions and are involved in DNA-related processes. A high proportion of multiple cellular localizations was predicted by the programs CELLO and Plant-mPloc. Nuc-Ploc suggested frequent nucleolar localization among the identified proteins (Fig. 6). These predictions, however, must be considered with caution and will require experimental evidence.

The UNcleProt database will be extended in the future by including tandem mass spectra (MS/MS spectra) for all sequenced peptides with their fragmentation patterns. Such a spectral library is expected to become useful as a reference for protein/peptide identification in other laboratories. The database is intended as an open system and will be continuously updated by adding newly discovered barley nuclear proteins. Links to all relevant publications and research output concerning the barley nuclear proteome will also be added.

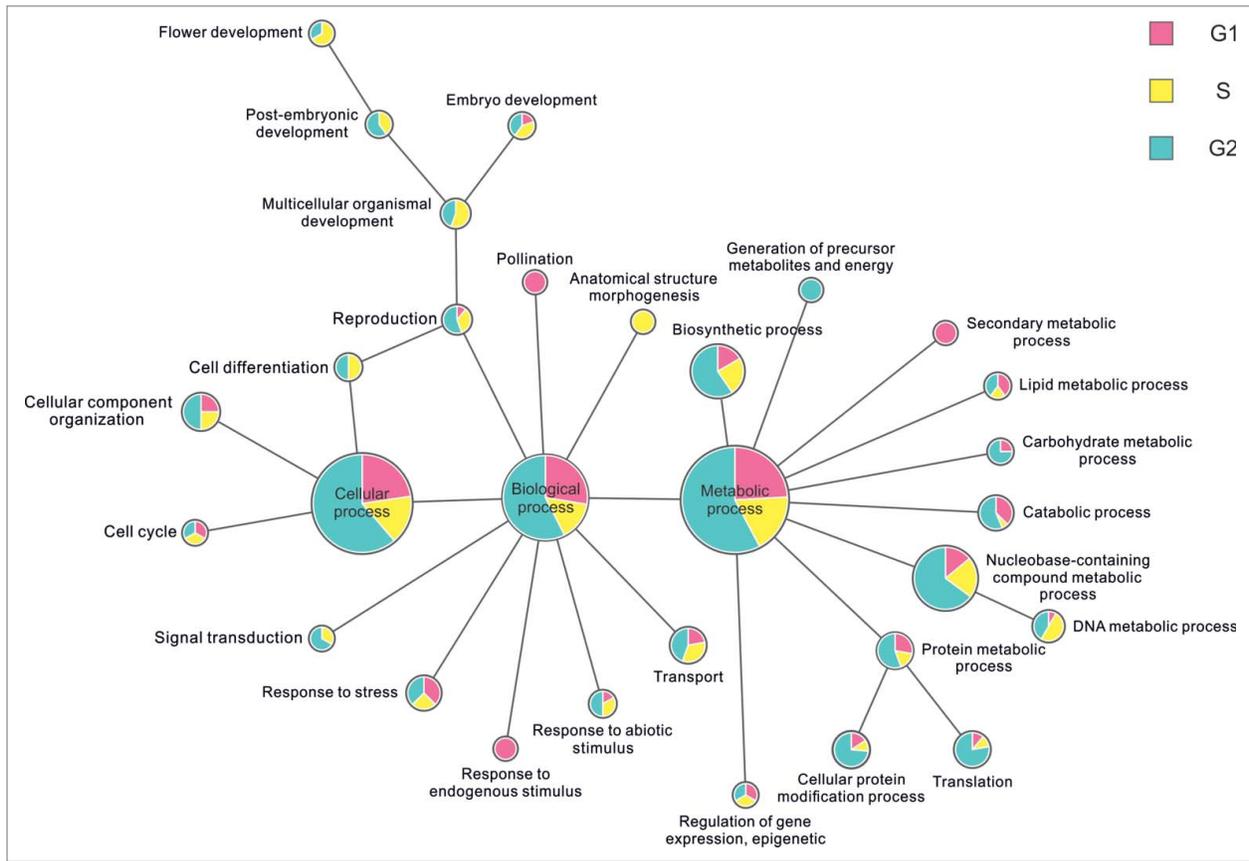


Figure 5. Extracted GO terms related to biological functions for proteins enriched in particular phases of the cell cycle. The radius of each circle denotes the total count belonging to the respective term and the pie chart inside shows its distribution across the cell cycle.

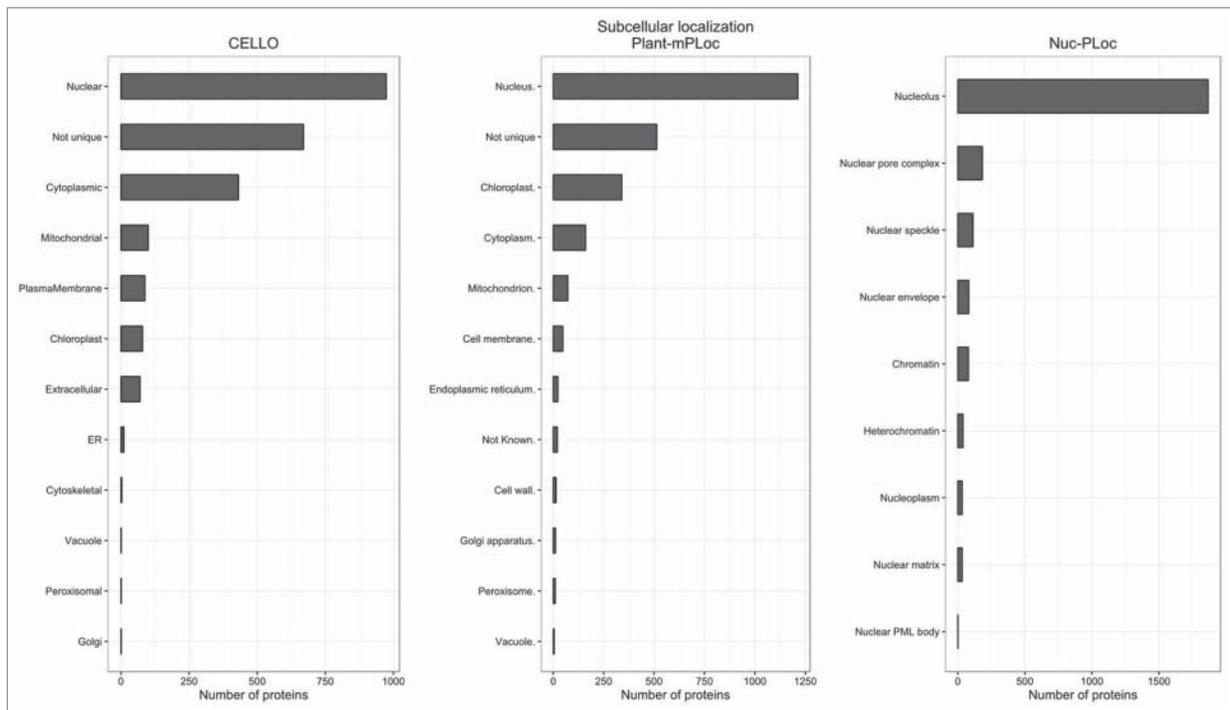


Figure 6. Predicted subcellular localization. The subcellular localization of nuclear proteins identified in 3 phases of the cell cycle as predicted by CELLO, Plant-mPLOC and Nuc-PLOC.

Materials and methods

The 6-rowed diploid malting cultivar of barley (*Hordeum vulgare* L.) cv Morex was used as an experimental plant material. All steps for the separation of barley nuclei and subsequent proteomic analyses were performed according to Petrovská et al.³³ Briefly, approx. 3 cm-long roots of barley were fixed in formaldehyde at 5°C for 10 min and washed twice in the Tris buffer. Root tips were homogenized in 1 ml of LB01-P lysis buffer (15 mM Tris, 2 mM Na₂EDTA, 0.5 mM EGTA, 80 mM KCl, 20 mM NaCl, 0.1% v/v Triton X-100, 0.2 mM spermine, 0.5 mM spermidine, 14 mM 2-mercaptoethanol). Crude homogenate was filtered through a 20- μ m nylon mesh, stained with 2 μ g/ml 4',6-diamidino-2-phenylindole (DAPI) and subjected to flow cytometric sorting (FACSARIA SORP, BD Biosciences, San Jose, Calif., USA). Nuclei at various phases of cell cycle (G1, S, G2) were sorted into tubes containing 1 ml of LB01-P buffer supplemented with 100 mM PMSF and pelleted (300 g, 4°C, 30 min).

Crude protein was extracted from the pelleted nuclei (5 million) together with a DNA digestion by DNase I in the treatment buffer.³³ The sample was centrifuged (25,000 g, 15 min) and proteins present in the supernatant were recovered by adding 4 volumes of cold acetone (−20°C) with incubation at −20°C for at least 24 h. The acetone precipitate was collected by centrifugation (25,000 g, 15 min) and finally dissolved in 50 μ l of Laemmli's sample buffer containing 2-mercaptoethanol, sonicated for 10 min and heated at 100°C for 10 min. The pellet obtained after the DNase I digestion was dissolved directly in Laemmli's buffer at 100°C.³³

Nuclear proteins were separated by SDS-PAGE in 4% stacking and 10% resolving gels and in-gel digested by a modified thermostable trypsin. The obtained peptides were purified and analyzed by liquid chromatography coupled with tandem mass spectrometry.³³

Raw MS and MS/MS data were processed by the software supplied with the instruments.³³ Database searches were performed using Mascot Server 2.4 search engine (Matrix Science, London, UK) against a custom-made barley protein sequence database (105,041 sequences) downloaded from the UniProt depository and supplemented with sequences of common contaminants and reversed sequences of the barley proteins for the determination of false discovery rate (FDR). The identified protein sequences were queried against *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays* subsets of the UniProtKB/Swiss-Prot database

using the BLASTP algorithm (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)³⁵ with BLOSUM62 matrix and following required thresholds: E-value of 1E-10; sequence identity of 70 %. For all assigned orthologs, protein names were downloaded from the UniProtKB/Swiss-Prot database. Domain searches were performed against CDD database⁵⁵ utilizing CD search tool in a batch mode (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). GO Retriever application provided by the AgBase database⁵⁶ was employed to predict GO terms for the respective identified proteins.

UNcleProt database is built on a MySQL v.5.5.46 (<http://www.mysql.com>) database, and the web interface uses PHP v.5.4.45 (<http://www.php.net>), which makes the database easy to navigate. UNcleProt is hosted on a Debian 7 server with 16 CPUs and 94 Gb (gigabit) of memory, located in the Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc. The database size is currently approximately 230 Mo (mega octet) and is expected to grow as new experimental data are added, with new database releases appearing accordingly. The input data to be processed and included are usually tab-delimited text files and any addition to the database can be possible by contacting the administrator (N. Blavet). Data downloading is easy, as described in the Results section.

The UNcleProt database is freely available at <http://barley.gambrinus.ueb.cas.cz/>. The website uses cookies to maintain information during the sessions and is compatible with all up-to-date major web browsers (tested on Firefox 46, Internet Explorer 11, Safari 5.34, Opera 37, Chrome 51).

Conclusions

UNcleProt is the first dedicated database containing plant nuclear proteins identified in nuclei during different stages of the cell cycle. This dataset will contribute to the understanding of plant nuclear proteins and their functions. It may become an important resource for plant cell biologists and contribute to the effort to understand the nuclear architecture and its relationship to genome function. Among other advantages for researchers, the database will facilitate the functional analyses of yet uncharacterized nuclear proteins.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

This work was supported by the Czech Science Foundation (grant award 14-28443S) and the Ministry of Education, Youth and Sports of the Czech Republic (National Program of Sustainability I award LO1204).

ORCID

Nicolas Blavet  <http://orcid.org/0000-0003-2199-9119>

Jan Vrána  <http://orcid.org/0000-0001-6171-8516>

Jaroslav Doležel  <http://orcid.org/0000-0002-6263-0492>

Beáta Petrovská  <http://orcid.org/0000-0002-0315-9872>

References

- [1] Guo T, Fang Y. Functional organization and dynamics of the cell nucleus. *Front Plant Sci* 2014; 5:378; PMID:25161658
- [2] Petrovská B, Šebela M, Doležel J. Inside a plant nucleus: discovering the proteins. *J Exp Bot* 2015; 66:1627-40; <http://dx.doi.org/10.1093/jxb/erv041>
- [3] Sutherland HGE, Mumford GK, Newton K, Ford LV, Farrall R, Dellaire G, Cáceres JF, Bickmore WA. Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Hum Mol Genet* 2001; 10:1995-2011; PMID:11555636; <http://dx.doi.org/10.1093/hmg/10.18.1995>
- [4] Wentz SR, Rout MP. The nuclear pore complex and nuclear transport. *Cold Spring Harb Perspect Biol* 2010; 2:a000562; PMID:20630994; <http://dx.doi.org/10.1101/cshperspect.a000562>
- [5] Aebi U, Cohn J, Buhle L, Gerace L. The nuclear lamina is a meshwork of intermediate-type filaments. *Nature* 1986; 323:560-4; PMID:3762708; <http://dx.doi.org/10.1038/323560a0>
- [6] Worman HJ, Gundersen GG. Here come the SUNs: a nucleocytoskeletal missing link. *Trends Cell Biol* 2006; 16:67-9; PMID:16406617; <http://dx.doi.org/10.1016/j.tcb.2005.12.006>
- [7] Graumann K, Bass HW, Parry G. SUNrises on the international plant nucleus consortium. *Nucleus* 2013; 4:3-7; PMID:23324458; <http://dx.doi.org/10.4161/nucl.23385>
- [8] Bae MS, Cho EJ, Choi EY, Park OK. Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *Plant J* 2003; 36:652-63; PMID:14617066; <http://dx.doi.org/10.1046/j.1365-313X.2003.01907.x>
- [9] Calikowski TT, Meulia T, Meier I. A proteomic study of the Arabidopsis nuclear matrix. *J Cell Biochem* 2003; 90:361-78; PMID:14505352; <http://dx.doi.org/10.1002/jcb.10624>
- [10] Bigeard J, Rayapuram N, Pflieger D, Hirt H. Phosphorylation-dependent regulation of plant chromatin and chromatin-associated proteins. *Proteomics* 2014; 14:2127-40; PMID:24889195; <http://dx.doi.org/10.1002/pmic.201400073>
- [11] Jones AME, MacLean D, Studholme DJ, Serna-Sanz A, Andreasson E, Rathjen JP, Peck SC. Phosphoproteomic analysis of nuclei-enriched fractions from Arabidopsis thaliana. *J Proteomics* 2009; 72:439-51; PMID:19245862; <http://dx.doi.org/10.1016/j.jprot.2009.02.004>
- [12] Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JWS, Shaw PJ. Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol Biol Cell* 2005; 16:260-9; PMID:15496452; <http://dx.doi.org/10.1091/mbc.E04-09-0791>
- [13] Aki T, Yanagisawa S. Application of rice nuclear proteome analysis to the identification of evolutionarily conserved and glucose-responsive nuclear proteins. *J Proteome Res* 2009; 8:3912-24; PMID:19621931; <http://dx.doi.org/10.1021/pr900187e>
- [14] Choudhary MK, Basu D, Datta A, Chakraborty N, Chakraborty S. Dehydration-responsive nuclear proteome of rice (*Oryza sativa* L.) illustrates protein network, novel regulators of cellular adaptation, and evolutionary perspective. *Mol Cell Proteomics* 2009; 8:1579-98; PMID:19321431; <http://dx.doi.org/10.1074/mcp.M800601-MCP200>
- [15] Jaiswal DK, Ray D, Choudhary MK, Subba P, Kumar A, Verma J, Kumar R, Datta A, Chakraborty S, Chakraborty N. Comparative proteomics of dehydration response in the rice nucleus: New insights into the molecular basis of genotype-specific adaptation. *Proteomics* 2013; 13:3478-97; PMID:24133045; <http://dx.doi.org/10.1002/pmic.201300284>
- [16] Khan MMK, Komatsu S. Rice proteomics: recent developments and analysis of nuclear proteins. *Phytochemistry* 2004; 65:1671-81; PMID:15276429; <http://dx.doi.org/10.1016/j.phytochem.2004.04.012>
- [17] Li G, Nallamilli BRR, Tan F, Peng Z. Removal of high-abundance proteins for nuclear subproteome studies in rice (*Oryza sativa*) endosperm. *Electrophoresis* 2008; 29:604-17; PMID:18203134; <http://dx.doi.org/10.1002/elps.200700412>
- [18] Mujahid H, Tan F, Zhang J, Nallamilli BRR, Pendarvis K, Peng Z. Nuclear proteome response to cell wall removal in rice (*Oryza sativa*). *Proteome Sci* 2013; 11:26; PMID:23777608; <http://dx.doi.org/10.1186/1477-5956-11-26>
- [19] Tan F, Li G, Chitteti BR, Peng Z. Proteome and phosphoproteome analysis of chromatin associated proteins in rice (*Oryza sativa*). *Proteomics* 2007; 7:4511-27; PMID:18022940; <http://dx.doi.org/10.1002/pmic.200700580>
- [20] Lee BJ, Kwon SJ, Kim SK, Kim KJ, Park CJ, Kim YJ, Park OK, Paek KH. Functional study of hot pepper 26S proteasome subunit RPN7 induced by Tobacco mosaic virus from nuclear proteome analysis. *Biochem Biophys Res Commun* 2006; 351:405-11; PMID:17070775; <http://dx.doi.org/10.1016/j.bbrc.2006.10.071>
- [21] Kumar R, Kumar A, Subba P, Gayali S, Barua P, Chakraborty S, Chakraborty N. Nuclear phosphoproteome of developing chickpea seedlings (*Cicer arietinum* L.) and protein-kinase interaction network. *J Proteomics* 2014;

- 105:58-73; PMID:24747304; <http://dx.doi.org/10.1016/j.jprot.2014.04.002>
- [22] Pandey A, Choudhary MK, Bhushan D, Chattopadhyay A, Chakraborty S, Datta A, Chakraborty N. The Nuclear proteome of chickpea (*Cicer arietinum* L.) reveals predicted and unexpected proteins. *J Proteome Res* 2006; 5:3301-11; PMID:17137331; <http://dx.doi.org/10.1021/pr060147a>
- [23] Pandey A, Chakraborty S, Datta A, Chakraborty N. Proteomics approach to identify dehydration responsive nuclear proteins from chickpea (*Cicer arietinum* L.). *Mol Cell Proteomics* 2008; 7:88-107; PMID:17921517; <http://dx.doi.org/10.1074/mcp.M700314-MCP200>
- [24] Subba P, Kumar R, Gayali S, Shekhar S, Parveen S, Pandey A, Datta A, Chakraborty S, Chakraborty N. Characterisation of the nuclear proteome of a dehydration-sensitive cultivar of chickpea and comparative proteomic analysis with a tolerant cultivar. *Proteomics* 2013; 13:1973-92; PMID:23798506; <http://dx.doi.org/10.1002/pmic.201200380>
- [25] Repetto O, Rogniaux H, Firnhaber C, Zuber H, Küster H, Larré C, Thompson R, Gallardo K. Exploring the nuclear proteome of *Medicago truncatula* at the switch towards seed filling. *Plant J* 2008; 56:398-410; PMID:18643982; <http://dx.doi.org/10.1111/j.1365-313X.2008.03610.x>
- [26] Casati P, Campi M, Chu F, Suzuki N, Maltby D, Guan S, Burlingame AL, Walbot V. Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. *Plant Cell* 2008; 20:827-42; PMID:18398050; <http://dx.doi.org/10.1105/tpc.107.056457>
- [27] Guo B, Chen Y, Li C, Wang T, Wang R, Wang B, Hu S, Du X, Xing H, Song X, et al. Maize (*Zea mays* L.) seedling leaf nuclear proteome and differentially expressed proteins between a hybrid and its parental lines. *Proteomics* 2014; 14:1071-87; PMID:24677780; <http://dx.doi.org/10.1002/pmic.201300147>
- [28] Abdalla KO, Rafudeen MS. Analysis of the nuclear proteome of the resurrection plant *Xerophyta viscosa* in response to dehydration stress using iTRAQ with 2D LC and tandem mass spectrometry. *J Proteomics* 2012; 75:2361-74; PMID:22361341; <http://dx.doi.org/10.1016/j.jprot.2012.02.006>
- [29] Abdalla KO, Baker B, Rafudeen MS. Proteomic analysis of nuclear proteins during dehydration of the resurrection plant *Xerophyta viscosa*. *Plant Growth Regul* 2010; 62:279-92; <http://dx.doi.org/10.1007/s10725-010-9497-2>
- [30] Cooper B, Campbell KB, Feng J, Garrett WM, Frederick R. Nuclear proteomic changes linked to soybean rust resistance. *Mol Biosyst* 2011; 7:773-83; PMID:21132161; <http://dx.doi.org/10.1039/C0MB00171F>
- [31] Bancel E, Bonnot T, Davanture M, Branlard G, Zivy M, Martre P. Proteomic approach to identify nuclear proteins in wheat grain. *J Proteome Res* 2015; 14:4432-9; PMID:26228564; <http://dx.doi.org/10.1021/acs.jproteome.5b00446>
- [32] Bonnot T, Bancel E, Chambon C, Boudet J, Branlard G, Martre P. Changes in the nuclear proteome of developing wheat (*Triticum aestivum* L.) grain. *Front Plant Sci* 2015; 6:905; PMID:26579155; <http://dx.doi.org/10.3389/fpls.2015.00905>
- [33] Petrovská B, Jerábková H, Chamrád I, Vrána J, Lenobel R, Urinová J, Šebela M, Doležel J. Proteomic analysis of barley cell nuclei purified by flow sorting. *Cytogenet Genome Res* 2014; 143:78-86; <http://dx.doi.org/10.1159/000365311>
- [34] International Barley Genome Sequencing Consortium A physical, genetic and functional sequence assembly of the barley genome. *Nature* 2012; 491:711-6; PMID:23075845
- [35] Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25:3389-402; PMID:9254694; <http://dx.doi.org/10.1093/nar/25.17.3389>
- [36] Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, et al. MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res* 2004; 32:D217-D222; PMID:14681398; <http://dx.doi.org/10.1093/nar/gkh095>
- [37] Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015; 43:D222-226; PMID:25414356; <http://dx.doi.org/10.1093/nar/gku1221>
- [38] Xiong E, Zheng C, Wu X, Wang W. Protein subcellular location: The gap between prediction and experimentation. *Plant Mol Biol Rep* 2015; 34:52-61; <http://dx.doi.org/10.1007/s11105-015-0898-2>
- [39] Chou KC, Shen HB. Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 2010; 5:e11335
- [40] Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins* 2006; 64:643-51; PMID:16752418; <http://dx.doi.org/10.1002/prot.21018>
- [41] Shen HB, Chou KC. Nuc-PLoc: a new web-server for predicting protein subnuclear localization by fusing PseAA composition and PsePSSM. *Protein Eng Des Sel* 2007; 20:561-7; PMID:17993650; <http://dx.doi.org/10.1093/protein/gzm057>
- [42] Lysák MA, Číhalíková J, Kubaláková M, Šimková H, Künzel G, Doležel J. Flow karyotyping and sorting of mitotic chromosomes of barley (*Hordeum vulgare* L.). *Chromosome Res* 1999; 7:431-44; <http://dx.doi.org/10.1023/A:1009293628638>
- [43] Vrána J, Kubaláková M, Šimková H, Číhalíková J, Lysák MA, Doležel J. Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics* 2000; 156:2033-41
- [44] Bartlett JG, Alves SC, Smedley M, Snape JW, Harwood WA. High-throughput Agrobacterium-mediated barley transformation. *Plant Methods* 2008; 4:22; PMID:18822125; <http://dx.doi.org/10.1186/1746-4811-4-22>

- [45] Kurowska M, Daszkowska-Golec A, Gruszka D, Marzec M, Szurman M, Szarejko I, Maluszynski M. TILLING - a shortcut in functional genomics. *J Appl Genet* 2011; 52:371-90; PMID:21912935; <http://dx.doi.org/10.1007/s13353-011-0061-1>
- [46] Bennett MD, Smith JB. Nuclear DNA Amounts in Angiosperms. *Philos Trans R Soc Lond, B, Biol Sci* 1976; 274:227-74; PMID:6977; <http://dx.doi.org/10.1098/rstb.1976.0044>
- [47] Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, et al. Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* 2011; 23:1249-63; PMID:21467582; <http://dx.doi.org/10.1105/tpc.110.082537>
- [48] Mayer KFX, Waugh R, Brown JWS, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, et al. A physical, genetic and functional sequence assembly of the barley genome. *Nature* 2012; 491:711-6; PMID:23075845
- [49] Luc PV, Tempst P. PINdb: a database of nuclear protein complexes from human and yeast. *Bioinformatics* 2004; 20:1413-5; PMID:15087322; <http://dx.doi.org/10.1093/bioinformatics/bth114>
- [50] Bickmore WA, Sutherland HGE. Addressing protein localization within the nucleus. *EMBO J* 2002; 21:1248-54; PMID:11889031; <http://dx.doi.org/10.1093/emboj/21.6.1248>
- [51] Dellaire G, Farrall R, Bickmore WA. The Nuclear Protein Database (NPD): sub-nuclear localisation and functional annotation of the nuclear proteome. *Nucleic Acids Res* 2003; 31:328-30; PMID:12520015; <http://dx.doi.org/10.1093/nar/gkg018>
- [52] Pijnappel WP, Kolkman A, Baltissen MP, Heck A, Jr, Timmers HM. Quantitative mass spectrometry of TATA binding protein-containing complexes and subunit phosphorylations during the cell cycle. *Proteome Sci* 2009; 7:46; PMID:20034391; <http://dx.doi.org/10.1186/1477-5956-7-46>
- [53] Tenga MJ, Lazar IM. Proteomic study reveals a functional network of cancer markers in the G1-Stage of the breast cancer cell cycle. *BMC Cancer* 2014; 14:710; PMID: 25252636; <http://dx.doi.org/10.1186/1471-2407-14-710>
- [54] Farkash-Amar S, Eden E, Cohen A, Geva-Zatorsky N, Cohen L, Milo R, Sigal A, Danon T, Alon U. Dynamic proteomics of human protein level and localization across the cell cycle. *PLoS One* 2012; 7:e48722; PMID:23144944; <http://dx.doi.org/10.1371/journal.pone.0048722>
- [55] Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015; 43:D222-226; PMID:25414356; <http://dx.doi.org/10.1093/nar/gku1221>
- [56] McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, et al. AgBase: a functional genomics resource for agriculture. *BMC Genomics* 2006; 7:229; PMID:16961921; <http://dx.doi.org/10.1186/1471-2164-7-229>