JPP JOURNAL OF PROTEINS AND PROTEOMICS
9(2), 2018, pp. 127-135

Research Article

NUCLEAR PROTEOME OF AN ORPHAN LEGUME - *CAJANUS CAJAN* L.

Pratigya Subba[1*], Sneha M. Pinto[1], Sandeep Kasaragod[1], Oishi Chatterjee[1,2,3], Mohd Altaf Najar[1] and T. S. Keshava Prasad[1,2*]

[1] *Center for Systems Biology and Molecular Medicine, Yenepoya Research Centre, Yenepoya (Deemed to be University), Mangalore 575018, India*
[2] *Institute of Bioinformatics, International Technology Park, Bangalore 560066, India*
[3] *School of Biotechnology, Amrita Vishwa Vidyapeetham, Kollam 690525, India*

Abstract: Nucleus, often referred to as the administrative center of the cell participates in the regulation of almost every major biological function. It occupies ~10% of the cell's volume. The nuclear proteome contributes to ~25% of the cellular proteome. Therefore, an understanding of the functional diversity of this organelle necessitates an in-depth characterization of its proteome component. In this study, we report the nuclear proteome of a grain legume pigeon pea using nuclear enrichment in conjunction with high-resolution mass spectrometry. Pigeon pea is the sixth most important legume crop grown mainly in the arid and semi-arid regions of the world. The enrichment of nuclei was verified by the identification various well characterized nuclear resident proteins. Proteins with regulatory functions such as kinases, phosphatases and transcription factors contributed to a significant percentage of the proteins identified. We also identified several proteins with no known biological functions. For these proteins, we performed bioinformatics analysis to determine their protein domain architecture. To the best of our knowledge, our data represents one of the largest catalogs of nuclear proteins from plant species reported till date.

Keywords: Orbitrap Fusion Tribrid mass spectrometer; histone; kinase; phosphatase.

Note: Coloured Figures and Supplementary Information are available on Journal Website in "Archives" Section

Introduction

Nuclei are ubiquitous organelles of eukaryotic cells that undertake major functions including storage of genetic material and regulation of gene expression. The roles of several plant nuclear proteins have been implicated in signal transduction networks activated under various biotic and abiotic stresses (Narula *et al.*, 2013). Apart from regulatory functions, nuclear proteome is also composed of structural proteins that contribute to the nucleoskeletal framework (Simon *et al.*, 2011). The organelle is bound by a phospholipid bilayer structure in which are embedded a vast number of integral membrane proteins. The outer membrane is in continuity with the rough endoplasmic membrane that serves as active sites for protein synthesis. These membranes are indented by a large number of pores that are composed of nucleoporin complexes. These pores facilitate the transportation of biological macro- and micromolecules between the nuclear matrix and cytoplasm (Wuhr *et al.*, 2015). The regulation of the entry and exit of these molecules offer distinct but dynamic composition to the nuclear matrix as compared to the cytoplasm. An added complexity is provided by the presence of the nucleolus within the nucleus. Known as the seat of ribosome assembly and synthesis, the role

Corresponding Author: **Thottethodi Subrahmanya Keshava Prasad. Pratigya Subba**
E-mail: keshav@yenepoya.edu.in; pratigyas@yenepoya.edu.in
Received: February 27, 2018
Accepted: June 28, 2018
Published: June 29, 2018

of nucleolus has also been implicated in various other cellular processes such as response to various stresses (Andersen *et al*., 2005).

Sub-cellular fractionation techniques offer several advantages to proteomics as they significantly reduce the complexity of the proteome being analyzed. The technique also allows the identification of low abundant regulatory proteins that are otherwise masked by abundant proteins. In conjunction with mass spectrometry, organelle enrichment techniques have allowed the identification of a large complement of proteins from various organelles (Bhushan *et al*., 2011; Bussell *et al*., 2013; Cui *et al*., 2011; Palm *et al*., 2016; Subba *et al*., 2013; Uberegui *et al*., 2015). Analysis of the nuclear proteome has been carried out in various plant species and differential regulation of nuclear proteins has also been studied under various conditions. In wheat, comparative nuclear proteome of developing seeds has been reported (Bancel *et al*., 2015; Bonnot *et al*., 2015). Dehydration-responsive nuclear proteome of 4-week-old seedlings have been reported in rice (Choudhary *et al*., 2009). Recently, nuclear proteome database of barley has been reported using nuclei isolated from different stages of the cell cycle (Blavet *et al*., 2017). Among dicots, nuclear proteome and phosphoproteome analysis have been conducted under flooding stress in soybean roots (Yin *et al*., 2015) whereas in *Arabidopsis* the analysis was conducted in response to chitosan treatment (Fakih *et al*., 2016).

Grain legumes (or pulses) serve as rich sources of proteins to the human body. Aptly referred to as poor man's meat, they also are sources of vitamins, minerals and dietary fibers. Apart from their nutritional value, they also have agricultural importance. Their symbiotic association with soil bacteria such as *Rhizobium* enables the conversion of atmospheric nitrogen to biologically usable form, ammonia. Globally, pigeon pea is the sixth most important grain legume (Varshney *et al*., 2011). Their deep tap-root system efficiently extracts soil water making them suitable for cultivation in the arid and semi-arid regions. The multifaceted functions of the nucleus qualify it to serve as a resource for the identification of target molecules that can be utilized for improvement of specific plant traits (Erhardt *et al*., 2010; Narula *et al*., 2013). In the current study, we isolated proteins from nuclear enriched fraction of pigeon pea seedlings. Enzymatic digestion was carried out using in gel and in solution methods. We identified 3826 proteins in toto. A large sub-set of proteins consisted of regulatory molecules such as kinases, phosphatases and transcription factors. The dataset will serve as an inventory for selecting target molecules for functional characterization that can be exploited for future crop improvement programs.

Materials and Methods

Plant growth, maintenance - Seeds of pigeon pea (*Cajanus cajan*) were purchased from the market and grown under field conditions (28±2 °C, 50% relative humidity). The plants were watered at regular intervals to maintain ~30% soil moisture content. The seedlings were observed on a daily basis to check for growth defects, if any. Two-week old seedlings were harvested and washed using running tap water. The harvested seedlings were then stored at -80 °C until further use.

Isolation of nuclei enriched fraction - Nuclei were isolated using the method described earlier with few modifications (Gendrel *et al*., 2005). The plant tissues (2 g) were ground to fine powder using mortar and pestle in liquid nitrogen and transferred to pre-chilled 50 ml centrifuge tubes. All procedures from here onwards were conducted on ice unless otherwise described. To the powder, 30 ml of extraction buffer 1 [0.4 M sucrose, 10 mM Tris-Cl (pH 8), 10 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride (PMSF)] was added and the tubes were inverted to mix the samples. The slurry was filtered through 2 layers of Miracloth (Calbiochem) and the filtrates were transferred to fresh 50 ml centrifuge tubes. The samples were centrifuged at 3,000 g for 20 min at 4 °C. The pellets were gently re-suspended in 1 ml of extraction buffer 2 [0.25 M sucrose, 10 mMTris-Cl (pH 8), 10 mM $MgCl_2$, 1 % Triton X-100, 5 mM 2-mercaptoethanol, 0.1 mM PMSF] and centrifuged at 12,000 g for 10 min at 4 °C. The pellets were re-suspended in 300 µl of extraction buffer 3 [1.7 M sucrose, 10 mM Tris-ClpH 8), 2 mM $MgCl_2$, 0.15 % Triton X-100, 5 mM 2-mercaptoethanol, 0.1 mM PMSF] and overlaid on to a fresh aliquot of 300 µl extraction buffer 3. The samples were centrifuged at 16,000 g for 1 h at 4 °C. The supernatants were discarded and the nuclear enriched pellets were re-suspended in 300 µl of 50 mM triethyl ammonium bicarbonate (TEABC) buffer.

Assessment of nuclear enrichment, protein extraction and trypsin digestion - An aliquot (20 µl) of the sample was stained using 4',6-Diamidine-2'-phenylindole dihydrochloride (DAPI). In brief, the sample was incubated with 0.1 µg/ml DAPI prepared in sterile water for 15 min and washed twice with 50 mM TEABC. For microscopy, 5 µl of the sample was placed on a glass slide and covered using a cover slip. Nuclei were viewed using a ZOE Fluorescent Cell Imager (Bio-Rad). Images were captured with or without a UV filter. Proteins were extracted from the nuclei enriched fraction using probe sonication followed by freeze-thaw cycles. To the isolated nuclei, 4% SDS lysis solution was added and incubated at room temperature for 15 min. The samples were then sonicated thrice and then snap-frozen in liquid nitrogen followed by thawing the sample at 37 °C. The freeze-thaw cycle were repeated 5 times. Finally, the samples were centrifuged at 12, 000 g for 30 min and the supernatants were transferred to fresh microcentrifuge tubes. Protein contents were estimated using bicinchoninic acid (BCA) assay.

Enzymatic digestion - Trypsin digestion of the proteins was carried out using in gel as well as in solution based methods. For in gel method, 100 µg of the proteins were separated using SDS-PAGE followed by staining using Coomassie Brilliant Blue. The protein gel lane was sliced into 12 gel bands that were further diced into smaller pieces (1mmX1mm). The samples were processed as described earlier (Kelkar *et al*., 2011). In brief, the gel pieces were de-stained followed by reduction using 5 mM dithiothreitol (DTT) and alkylation using 20 mMiodoacetamide (IAA). Trypsin (Promega, Madison, WI) digestion was carried out overnight at 37 °C. The tryptic peptides were transferred to fresh microcentrifuge tubes and dried using a SpeedVac Concentrator (Thermo Scientific). The peptides were de-salted using C_{18} Stage Tips.

For in solution digestion, 100 µg of proteins were subjected to reduction using 10 mM DTT and alkylation using 20 mM IAA. Trypsin (Worthington Biochemical Corp) digestion was carried at 37 °C. Fractionation of peptides was carried out using basic reversed phase liquid chromatography (bRPLC). Using a manual injector, the peptide digests were injected onto XBridge C18 column (5 µm 250 x 4.6 mm; Waters, Milford, MA) connected to LaChrom Elite HPLC system (Hitachi). The gradient used for the fractionation was 0 to 100% solvent B (10mM TEABC in acetonitrile, pH 8.5) resolved in 130 minutes. The flow rate was set to 0.5 ml/min. 96 fractions collected were finally concatenated into 12 fractions, dried and desalted using C_{18} Stage Tips.

LC-MS/MS analysis - The dried peptides were re-constituted in 0.1% formic acid and analyzed using Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific) interfaced with Easy-nLC II nanoflow liquid chromatography system (Thermo Scientific, Odense, Denmark). Peptides were loaded onto nanoViper trap column (2 cm, 3 um, C18 Aq) (Thermo Fisher Scientific) and then resolved using a nanoViper analytical column (15 cm, 75 um silica capillary, 2 um C18 Aq) (Thermo Fisher Scientific). The solvents used for the separation of peptides were solvent A: 0.1% formic acid in water and solvent B: 0.1% formic acid in 80% acetonitrile. The solvent gradients were set as follows: linear gradient of 5-35% solvent B over 100 minutes followed by 100% solvent B for 15 min and 95% solvent A for 5 min for a total run time of 120 mins. The flow rate was set to 250 nl/min. Data were acquired in a positive scan mode. Data-dependent acquisition was carried out using MS1 survey scans in 400-1600 m/z range (120,000 mass resolution at 400 m/z). Subsequently, for every MS scan the most intense precursor ions were selected for MS/MS analysis at top speed data dependent mode with maximum cycle time of 3 seconds. The parameters were set as follows: HCD fragmentation, collision energy 33%, mass resolution 30,000. Peptides with charge 2-6 were selected, the dynamic exclusion was set to 30 sec and exclusion width of ± 20 ppm. Internal calibration was carried out using lock mass option (m/z 445.1200025) from ambient air. Peptides were identified by searching the raw files against the *Cajanus cajan* protein database downloaded from NCBI. Common protein contaminants were appended to the database and searches were performed using SEQUEST and MASCOT algorithms through Proteome Discoverer platform suite version 2.1; (Thermo Scientific). For both the algorithms, search parameters included maximum of two missed cleavages, peptide tolerance, 10 ppm and fragment mass tolerance, 0.05 Da. A false discovery rate (FDR) was enabled using a decoy database wherein a cut-off of 1% peptide spectral match and 1% peptide level FDR were used for the analysis.

Protein domain analysis - Protein sequences for the protein descriptions associated with the following terms: 'hypothetical protein',

'uncharacterized protein' or associated with UPF numbers were subjected to protein domain analysis. The analysis was performed using the SMART domain analysis tool (http://smart.embl-heidelberg.de/) (Letunic *et al*., 2012).

Results and Discussion

Subcellular fractionation for enrichment of nuclei and purity assessment

The identification of proteins from specific sub-cellular compartments is important to understand the functions carried out by these molecules. Sub-cellular fractionation techniques also aid in the identification of proteins present in very low abundance as, these techniques significantly reduce the complexity of the proteome. We isolated enriched nuclear fractions using a differential centrifugation based method and proteins were extracted for mass spectrometric analysis. An overview of the experimental workflow carried out is depicted in Figure 1. The purity assessment and intactness of the isolated nuclei was conducted using DAPI, which stains double stranded DNA strands (Figure 2). Photographs were taken using fluorescence microscopy. Spherical nuclei with sizes approximately 10 μm in diameter were observed. The nuclear proteins were isolated using 4 % SDS solution in conjunction with sonication and repeated cycles of freeze-thaw in liquid nitrogen. The enrichment of nuclear fraction was confirmed by the identification of several well-known nuclear resident proteins such as Histone (H1, H1.2, H2A, H2AX, H2B 7, H3, H3.1, H3.2, H4), HMG1/2-like protein, histone acetyltransferase, several variants of histone deacetylases, small nuclear ribonucleoproteins, nuclear pore complex proteins (Nup53, Nup133 family, Nup155 family, Nup160 family, Nup205 family, Nuclear pore complex protein Nup98-Nup96) and zinc finger CCCH domain-containing proteins among others.

Mass spectrometric identification of nuclear proteins

Samples were processed using two methods of trypsinization *viz*., in gel (SDS PAGE) and in solution prior to mass spectrometric analysis. The SDS PAGE lanes containing the proteins were divided into 12 bands whereas in solution digested tryptic peptides were fractionated into 12 fractions using bRPLC. From the mass spectrometric analysis, a total of 396,329 MS/MS spectra were acquired which could be assigned to 24,939 peptide sequences using the Sequest and Mascot search algorithms. Only proteins identified with at least two peptides were considered for further analysis. In total, the analysis led to the identification of 3826 protein groups. Of these, 924 proteins were identified exclusively in the bRPLC method whereas 450 proteins were identified only in the in gel method. A venn diagram depicting the common and unique proteins is shown in Figure 3A. All the proteins identified in this study along with the details such as scores, charge, number of unique peptides, PSMs etc. are enlisted in Supplementary Table 1, 2. The raw mass spectrometry data (.raw format) have been submitted to the PRIDE database using the link provided by PRIDE (PXD008732).

Classification of the identified proteins

The identified proteins were classified based upon their physicochemical characteristics. The molecular weights range from ~5-500 Da whereas the pIs range from pH 4 to pH 12 (Figure 3B, 3C). The identified proteins were also categorized according to their biological functions (Figure 3D, 3E). The functional classes were assigned based upon the output of the Proteome Discoverer software. The highest percentage of proteins belonged to the class 'metabolic process' followed by classes 'transport' and 'cell organization and biogenesis'. In the class 'response to stimulus' we identified 78 proteins. The outer membranes of nuclei are in continuation with rough endoplasmic reticulum that serves as sites for protein synthesis. Ribosomal proteins are some of the most abundant proteins in a cell, accounting for nearly 40% of the proteome. A large number of ribosomal 40S and 60S sub-units were identified in this study. We identified 127 proteins associated with the term kinase. The list includes kinases such as TP53-regulating kinase-like, dual specificity protein kinase splB-like and SNF1-related protein kinase catalytic subunit. The human TP53-regulating kinase-like proteinis known to phosphorylate the tumor suppressor protein p53 at Ser15 (Abe *et al*., 2001). However, in *A. thaliana* the protein has not been associated with any known function thereby making it an interesting candidate for functional characterization. The *A. thaliana* SNF1-related protein kinase has been reported to be activated under osmotic stresses and abscisic acid-regulated plant development (Bucholc *et al*., 2011; Fujii *et al*., 2011). The protein has been reported to be negatively regulated by the ABI1 and PP2CA

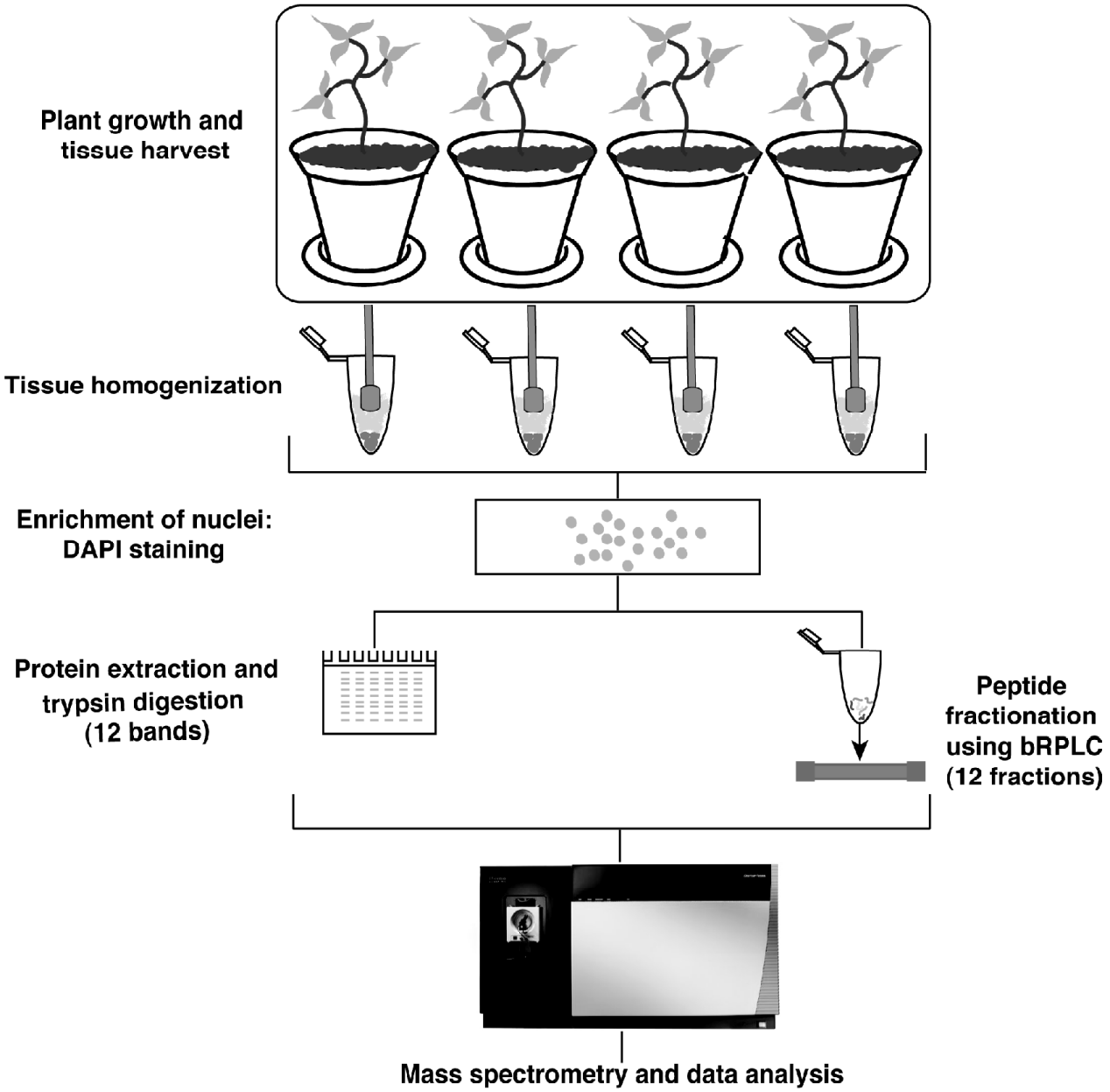

Figure 1: **Workflow depicting the experimental procedure**

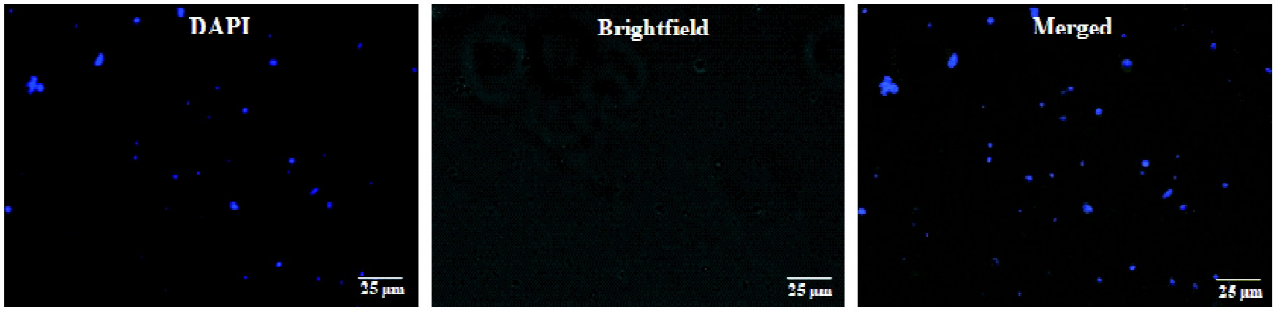

Figure 2: **Purity assessment of the isolated nuclei. The intactness and enrichment of nuclei were assessed using DAPI (left panel) and visualized using fluorescence microscopy. The corresponding brightfield and merged images are shown in the right panels**

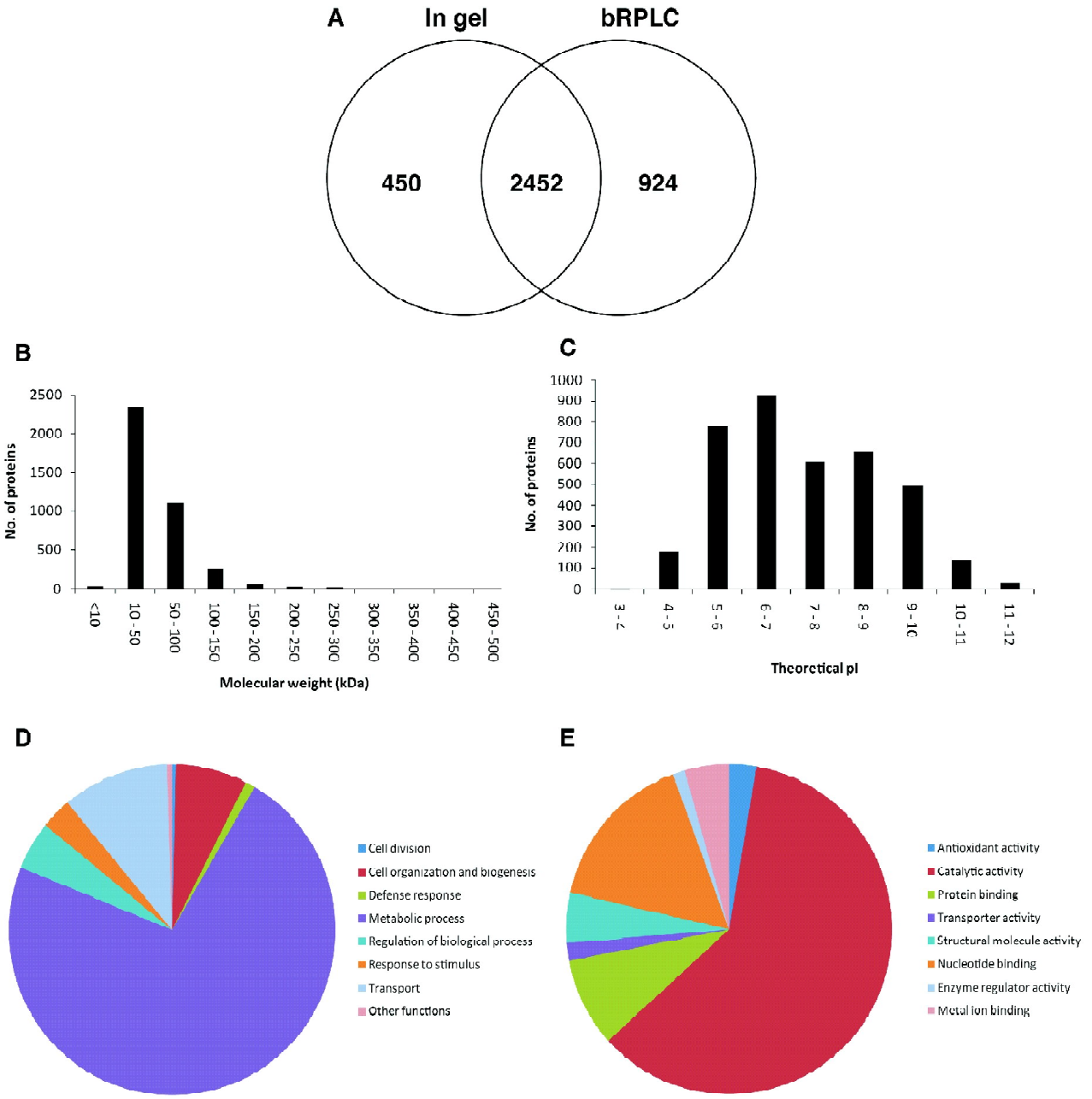

Figure 3: **Overlap of proteins identified using two fractionation techniques. (A) Venn diagram depicts the common and unique proteins identified in the bRPLC and in gel digestion methods. (B) Distribution of the molecular weights. (C) Distribution of the iso- electric point pIs. Classification of the identified proteins based on (D) biological process and (E) molecular function**

Phosphatases (Rodrigues *et al.*, 2013). Interestingly, the sub-cellular localization for the *A. thaliana* homologs of SNF1-related protein kinase and PP2CA Phosphatase has been reported as nucleus, cytosol and membrane on the TAIR database. The co-localization of these proteins thereby validates their functional dependency. As these proteins are known to be regulated under several stresses, in future SNF1-related protein kinase and PP2CA Phosphatase identified in our study can be exploited to generate multi-stress tolerant crop plants. Our list of identified proteins include 62 phosphatases including Serine/threonine-protein phosphatase BSL1, serine/threonine protein phosphatase 2A 59 kDa, protein-tyrosine-phosphatase MKP1 and Dual specificity

phosphatase Cdc25. Although phosphorylation of tyrosine residues under various conditions has been reported in plants, no plant Tyr kinase has been reported. The regulation is carried out by dual specific kinases and phosphatases. Expression of *S. pombe* cdc25 in BY-2 cells perturbed the cytokinin levels and promoted mitosis and root production (Spadafora *et al*., 2012). Functional roles of cdc25 with respect to plant growth and development and hormone signaling can be studied in future. The Serine/threonine-protein phosphatase BSL1 is a BSU1-like protein 1that participates in brassinolide signaling. In *A. thaliana* BSU1 dephosphorylates the kinase BES1 (phosphorylated by BIN2 GSK3 kinase) that leads to the activation of brassinosteroid signaling. The WoLF PSORT tool (https://wolfpsort.hgc.jp/) predicted nuclear localization for the protein Putative Holliday junction resolvase. We identified 7 proteins that shared homology with various *A. thaliana* putative inactive receptor kinases. We also identified 2 putative WD repeat-containing proteins. WD40 domain-containing proteins contain several repeats of the amino acids tryptophan and aspartic acid. Functional roles for these domains have been implicated in various biological processes through protein-protein interactions (van Nocker and Ludwig, 2003). Proteins containing basic helix-loop-helix (bHLH) DNA binding domains have been reported to be localized in the nucleus in *A. thaliana*. Roles of these transcription factors have been implicated in diverse biological processes such as plant development and light signaling (Castillon *et al*., 2007; Ramsay and Glover, 2005). Legume specific stress regulatory functions of bHLH proteins can be studied as legumes are known to be naturally resilient to stresses such as water-deficit stress. All together, we identified 12 proteins associated with the term 'transcription factor'. The list included transcription factor MYB39-like, ethylene-responsive transcription factor, nuclear transcription factor Y subunit C-1 and bZIP transcription factor TGA10-like, transcript. We identified several (24) subunits of the proteasome complex such as 26S proteasome non-ATPase regulatory subunit 12, partial, Proteasome subunit alpha type-3 and Proteasome subunit beta type-6. Identification of nuclear resident proteins such as Histones, Histone deacetylase HDT1, Nuclear pore complex protein family, Nuclear pore glycoprotein p62 and HMG1/2-like protein validated the enrichment of nuclei in the preparation.

Protein domain analysis

Several proteins (1093) were identified with no known biological functions These proteins were associated with either of the following terms 'Hypothetical protein', 'Uncharacterized protein' or possessed 'UPF' numbers. In order to predict their putative functions, these protein sequences were subjected to protein domain analysis using SMART search engine. The submission of sequences was done using programmatic batch submission option. The number of domains predicted in each of the proteins ranged from 1 to 207 (Supplementary Table S3). Among the domains identified, 423 Domains of unknown function (DUF) were found in various proteins. Together with UPF domain containing proteins, they form the DUF family of proteins (Bateman *et al*., 2010; Mudgal *et al*., 2015). Other predicted domains include 59 kinase domains, 12 Cullin domains and 44 leucine-rich repeats (LRR). This analysis is the first functional annotation assigned to these proteins. As these proteins accounted for a significant percentage (~28%) of the nuclear proteome dataset, it will be interesting to explore their functions in various legumes.

Nuclear proteomics of plants

We catalogued all literature available on plant nuclear proteomic studies till date. The data including research articles, reviews and methods for isolation of nucleus and extraction of nuclear proteins were retrieved using Pubmed using the search terms 'Plant' AND 'Nucleus' AND 'Proteomics' (Supplementary Table S4). The studies were conducted either to report the reference proteome maps or to understand the differential regulation under various conditions such as water-deficit, flooding stress and seed development among others. Among monocotyledons, nuclear proteomics has been conducted in *Oryza sativa*, *Zea mays* and *Triticum aestivum*, *Xerophyta viscose*. The starting materials in *O. sativa* include both, seedlings as well as suspension cell cultures (Choudhary *et al*., 2009; Tan *et al*., 2007). Both, gel-based as well as gel-free proteomics have been utilized for analyzing the nuclear proteomes. It may be noted that we have also included proteomic studies conducted using nucleolus (Brown *et al*., 2005; Palm *et al*., 2016; Pendle *et al*., 2005). Among dicotyledons, nuclear proteome has been analyzed in *A. thaliana*, *Solanum lycopersicum, Glycine max, Medicago truncatula* and *Cicer arietinum*. The highest number of nuclear

proteins has been reported in *G. max* under *Phakopsora pachyrhizi* infection (Cooper *et al*., 2011). The list also includes two databases containing the catalog of nuclear proteins in *Hordeum vulgare* and *A. thaliana* (Blavet *et al*., 2017; Brown *et al*., 2005).

Conclusions

Organellar proteomics can be exploited for the identification of proteins that are localized to a particular sub-cellular compartment. This technique also allows the identification of low abundant proteins which are often masked by the highly abundant proteins. In the current study, we carried out the nuclear proteome analysis of a grain legume - pigeon pea using high-resolution mass spectrometry. We identified a number of well-characterized nuclear resident proteins validating the enrichment of nuclei preparation. A number of proteins were also identified with no known functions. Protein domain analysis of these proteins was carried out to associate them with putative biological functions. We report here one of the largest repertoire of nuclear proteins carried out in a plant species.

Acknowledgments

We thank Karnataka Biotechnology and Information Technology Services (KBITS), Government of Karnataka for the grant support to the Center for Systems Biology and Molecular Medicine at Yenepoya (Deemed to be University) under the Biotechnology Skill Enhancement Programme (BiSEP) in Multiomics Technology. PS is funded by the Early Career Research Award (ECR/2016/000365) from Science & Engineering Research Board (SERB), Government of India. SK is a Junior Research Fellow funded by Yenepoya (Deemed to be University). SMP is a recepient of DST-INSPIRE Faculty award (LSBM-138/2015) from the Department of Science &Technology (DST), Government of India. OC is a recipient of INSPIRE Fellowship from the Department of Science and Technology (DST), Government of India. Mohd Altaf Najar is a recipient of Junior Research Fellowship from University Grants Commission (UGC), Government of India. We acknowledge the Yenepoya (Deemed to be University) for access to mass spectrometry instrumentation facility.

Conflict of Interest

No competing financial interests exist.

Abbreviations

BCA, Bicinchoninic Acid Assay; bRPLC, Basic reverse phase liquid chromatography; HCD, Higher-energy collision dissociation; HPLC, High-performance liquid chromatography; LC-MS/MS, Liquid chromatography tandem-mass spectrometry; PSM, Peptide-spectrum match; SDS–PAGE, Sodium dodecyl sulfate polyacrylamide gel electrophoresis; TEABC, Triethyl ammonium bicarbonate.

References

Abe Y, Matsumoto S, Wei S, *et al*. (2001). Cloning and characterization of a p53-related protein kinase expressed in interleukin-2-activated cytotoxic T-cells, epithelial tumor cell lines, and the testes. J Biol Chem 276, 44003-44011.

Andersen JS, Lam YW, Leung AK, *et al*. (2005). Nucleolar proteome dynamics. Nature 433, 77-83.

Bancel E, Bonnot T, Davanture M, Branlard G, Zivy M and Martre P (2015). Proteomic Approach to Identify Nuclear Proteins in Wheat Grain. J Proteome Res 14, 4432-4439.

Bateman A, Coggill P and Finn RD (2010). DUFs: families in search of function. Acta Crystallogr Sect F Struct Biol Cryst Commun 66, 1148-1152.

Bhushan D, Jaiswal DK, Ray D, *et al*. (2011). Dehydration-responsive reversible and irreversible changes in the extracellular matrix: comparative proteomics of chickpea genotypes with contrasting tolerance. J Proteome Res 10, 2027-2046.

Blavet N, Urinovska J, Jerabkova H, *et al*. (2017). UNcleProt (Universal Nuclear Protein database of barley): The first nuclear protein database that distinguishes proteins from different phases of the cell cycle. Nucleus 8, 70-80.

Bonnot T, Bancel E, Chambon C, Boudet J, Branlard G and Martre P (2015). Changes in the nuclear proteome of developing wheat (Triticum aestivum L.) grain. Front Plant Sci 6, 905.

Brown JW, Shaw PJ, Shaw P and Marshall DF (2005). Arabidopsis nucleolar protein database (AtNoPDB). Nucleic Acids Res 33, D633-636.

Bucholc M, Ciesielski A, Goch G, *et al*. (2011). SNF1-related protein kinases 2 are negatively regulated by a plant-specific calcium sensor. J Biol Chem 286, 3429-3441.

Bussell JD, Behrens C, Ecke W and Eubel H (2013). Arabidopsis peroxisome proteomics. Front Plant Sci 4, 101.

Castillon A, Shen H and Huq E (2007). Phytochrome Interacting Factors: central players in phytochrome-mediated light signaling networks. Trends Plant Sci 12, 514-521.

Choudhary MK, Basu D, Datta A, Chakraborty N and Chakraborty S (2009). Dehydration-responsive nuclear proteome of rice (Oryza sativa L.) illustrates protein network, novel regulators of cellular adaptation, and evolutionary perspective. Mol Cell Proteomics 8, 1579-1598.

Cooper B, Campbell KB, Feng J, Garrett WM and Frederick R (2011). Nuclear proteomic changes linked to soybean rust resistance. Mol Biosyst 7, 773-783.

Cui J, Liu J, Li Y and Shi T (2011). Integrative identification of Arabidopsis mitochondrial proteome and its function exploitation through protein interaction network. PLoS One 6, e16022.

Erhardt M, Adamska I and Franco OL (2010). Plant nuclear proteomics—inside the cell maestro. FEBS J 277, 3295-3307.

Fakih Z, Ahmed MB, Letanneur C and Germain H (2016). An unbiased nuclear proteomics approach reveals novel nuclear protein components that participates in MAMP-triggered immunity. Plant Signal Behav 11, e1183087.

Fujii H, Verslues PE and Zhu JK (2011). Arabidopsis decuple mutant reveals the importance of SnRK2 kinases in osmotic stress responses in vivo. Proc Natl Acad Sci U S A 108, 1717-1722.

Gendrel AV, Lippman Z, Martienssen R and Colot V (2005). Profiling histone modification patterns in plants using genomic tiling microarrays. Nat Methods 2, 213-218.

Kelkar DS, Kumar D, Kumar P, *et al.* (2011). Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry. Mol Cell Proteomics 10, M111 011627.

Letunic I, Doerks T and Bork P (2012). SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40, D302-305.

Mudgal R, Sandhya S, Chandra N and Srinivasan N (2015). De-DUFing the DUFs: Deciphering distant evolutionary relationships of Domains of Unknown Function using sensitive homology detection methods. Biol Direct 10, 38.

Narula K, Datta A, Chakraborty N and Chakraborty S (2013). Comparative analyses of nuclear proteome: extending its function. Front Plant Sci 4, 100.

Palm D, Simm S, Darm K, *et al.* (2016). Proteome distribution between nucleoplasm and nucleolus and its relation to ribosome biogenesis in Arabidopsis thaliana. RNA Biol 13, 441-454.

Pendle AF, Clark GP, Boon R, *et al*. (2005). Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol Biol Cell 16, 260-269.

Ramsay NA and Glover BJ (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci 10, 63-70.

Rodrigues A, Adamo M, Crozet P, *et al.* (2013). ABI1 and PP2CA phosphatases are negative regulators of Snf1-related protein kinase1 signaling in Arabidopsis. Plant Cell 25, 3871-3884.

Simon DN and Wilson KL (2011). The nucleoskeleton as a genome-associated dynamic 'network of networks'. Nat Rev Mol Cell Biol 12, 695-708.

Spadafora ND, Parfitt D, Marchbank A, *et al*. (2012). Perturbation of cytokinin and ethylene-signalling pathways explain the strong rooting phenotype exhibited by Arabidopsis expressing the Schizosaccharomyces pombe mitotic inducer, cdc25. BMC Plant Biol 12, 45.

Subba P, Kumar R, Gayali S, *et al.* (2013). Characterisation of the nuclear proteome of a dehydration-sensitive cultivar of chickpea and comparative proteomic analysis with a tolerant cultivar. Proteomics 13, 1973-1992.

Tan F, Li G, Chitteti BR and Peng Z (2007). Proteome and phosphoproteome analysis of chromatin associated proteins in rice (Oryza sativa). Proteomics 7, 4511-4527.

Uberegui E, Hall M, Lorenzo O, Schroder WP and Balsera M (2015). An Arabidopsis soluble chloroplast proteomic analysis reveals the participation of the Executer pathway in response to increased light conditions. J Exp Bot 66, 2067-2077.

van Nocker S and Ludwig P (2003). The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. BMC Genomics 4, 50.

Varshney RK, Chen W, Li Y, *et al*. (2011). Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30, 83-89.

Wuhr M, Guttler T, Peshkin L, *et al*. (2015). The Nuclear Proteome of a Vertebrate. Curr Biol 25, 2663-2671.

Yin X and Komatsu S (2015). Quantitative proteomics of nuclear phosphoproteins in the root tip of soybean during the initial stages of flooding stress. J Proteomics 119, 183-195.

Supplementary Files

Supplementary Table S1: Summary of proteins identified using LC-MS/MS analysis

Supplementary Table S2: Summary of peptides identified using LC-MS/MS analysis

Supplementary Table S3: Details regarding the conserved proteins domain of unidentified the nuclear proteins

Supplementary Table S4: A summary of scientific literature available on plant nuclear proteomics studies