Dataset Brief

Proteome profile of the endomembrane of developing coleoptiles from switchgrass (*Panicum virgatum*)

Jeemeng Lao[1], Manoj Sharma[1,2,3], Rita Sharma[1,2,4], Susana M. González Fernández-Niño[1], Jeremy Schmutz[5,6], Pamela C. Ronald[1,2], Joshua L. Heazlewood[1,5] and Benjamin Schwessinger[1,2]

[1]Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[2]Department of Plant Pathology, University of California, Davis, CA, USA

[3]School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India

[4]School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

[5]Department of Energy, Joint Genome Institute, Walnut Creek, CA, USA

[6]HudsonAlpha Genome Sequencing Center, Huntsville, AL, USA

[7]ARC Centre of Excellence in Plant Cell Walls, School of Botany, The University of Melbourne, Victoria, Australia.

Corresponding Author:
Benjamin Schwessinger
Department of Plant Pathology
University of California,
Davis, California 95616
United States of America
Telephone:
E-mail: bschwessinger@ucdavis.edu

Keywords

Switchgrass, cell wall, endomembrane, coleoptile, proteome

Abstract

The cost-effective production of biofuels from lignocellulosic material will likely require manipulation of plant biomass, specifically cell walls. The North American native prairie grass *Panicum virgatum* (switchgrass) is currently seen as a potential biofuel crop with an array of genetic resources currently being developed. We have characterized the endomembrane proteome of switchgrass coleoptiles to provide additional information to the switchgrass community. In total, we identified 1,750 unique proteins from two biological replicates. The data have been deposited in the ProteomeXchange with the identifier PXD001351.

Plant cell walls are highly variable in content but essentially comprise the polysaccharides cellulose, hemicellulose and pectin with secondary cell walls also containing the cross-linked phenolic polymer lignin. Cellulose is comprised of glucose molecules and is synthesized by the cellulose synthase complex on the plasma membrane [1]. The matrix polysaccharides (hemicellose and pectin) are synthesized in the Golgi apparatus from a range of C_5 and C_6 sugars and secreted to the cell wall forming a flexible cross-linking matrix [2, 3]. Plant cell walls or biomass is costly to deconstruct for sugar release for downstream applications such as biofuel production. Nonetheless, modifications to the structure could result in significant downstream economic impacts [4]. The plant endomembrane comprises membrane systems from core organelles involved in cell wall biosynthesis, including the ER, the Golgi apparatus and the plasma membrane. Basic functional information gleaned from their proteomes will be required to fully exploit plant biomass for the development of cost effective biofuels [5].

The genetic manipulation of plant species to produce a tailored biomass for biofuels production will necessitate the development of genetic tools and resources. For the past two decades a concerted effort has been made to develop such resources for the perennial grass, *Panicum virgatum* (switchgrass), a native to North America [6]. A coordinated effort is currently underway to sequence the tetraploid Alamo AP13 clones of switchgrass, using next generation sequencing platforms [7]. Various genomic resources including large-insert physical clones, dense genetic maps and whole genome sequence of diploid switchgrass accessions have been developed to assists the assembly and annotation [6]. The current genome assembly for switchgrass is version 1.1 comprising more than 300,000 contigs and is available through Phytozome [8]. Proteomic analyses are a valuable resource for validating gene predictions, refining predicted gene models and identify novel genes missing from the

current annotation [9]. In the case of polyploid and heterozygous genomes like switchgrass, proteomics has the potential to correctly identify the functional alleles.

In an effort to make use of the developing genome resources and to provide candidates to the community that could be targeted for genetic manipulation, we sought to profile the proteome of the switchgrass endomembrane from 10-day old dark grown coleoptiles. Around 200 mL of switchgrass seeds were washed and aerated overnight at room temperature. The seeds were spread into moistened coarse-grade vermiculite and left to incubate for a week in the dark at 30°C. The harvested coleoptiles (40 g) were added to 80 mL of charcoal grinding buffer (200 mM HEPES, 40 mM KCl, 20 mM Sodium ascorbate, adjusted to pH 7.6 with 1 M Bis-tris propane, 84 % w/v sucrose and 12 % w/v of 100 mesh charcoal) and left for 10 min to reduce turgor [10]. The coleoptiles were ground in a mortar and pestle and homogenate filtered through miracloth. The density of the homogenate was adjusted with water or sucrose to be between the density of 35 and 46 % sucrose. A sucrose gradient comprising 8 mL of 35 % sucrose, 8 mL of 29 % sucrose, and 2.5 mL of 18.5 % sucrose in gradient buffer (10 mM HEPES and 20 mM KCl, adjusted to pH 7.2 with 1 M Bis-tris propane) was placed on top of the homogenate and the sample centrifuged at 100,000 g at 4 °C for 90 min in a swing bucket rotor. The band corresponding to the endomembrane fraction at the 29 % and 35 % interface was collected and protein concentration estimated by Bradford [11]. The degree of enrichment of endomembrane proteins from switchgrass coleoptiles was assessed by immunoblotting using antibodies against organellar marker proteins (Figure S1). Both plasma membrane and ER markers (endomembrane) showed considerable enrichment, with cytosolic, nuclear, plastid and mitochondrial markers depleted from the endomembrane fraction. A total of 5 μg of the endomembrane sample was digested in a solution of Tris-HCl (pH 7.5) and 40% methanol with trypsin (1:10 w/w) overnight at 37°C and purified according

to previous approaches [12]. Endomembrane fractions from two biological replicates were prepared and each fraction was analyzed twice by LC-MS/MS (Figure 1).

Digested peptides (ca. 1 ug) were analyzed by LC-MS/MS on a Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific) using a Proxeon Easy-nLC II HPLC (Thermo Scientific) and Proxeon nanospray source. The digested peptides were loaded on a 100µm x 25 mm Magic C18 100Å 5µm (Michrom) reverse phase trap for online desalting before separation on a 75 µm x 150 mm Magic C18 200Å 3µm (Michrom) reverse phase column. Peptides were eluted using a 90 minute gradient with a flow rate of 300 nl/min. An MS survey scan was obtained for the m/z range 300 to 1600. MS/MS spectra were acquired using a top 15 method, where the top 15 ions in the MS spectra were subjected to HCD (High Energy Collisional Dissociation). An isolation mass window of 2.0 m/z was used for the precursor ion selection, and normalized collision energy of 27% was used for fragmentation. A duration of 5 s was used for the dynamic exclusion. An Automatic Gain Control (AGC) target of 1000000 for MS and 50000 for MS/MS was used, while maximum IT for MS was 30 ms and MS/MS was 50 ms. The system employed a resolution of 70,000 for MS and 17,500 for MS/MS. Tandem mass spectra were extracted, charge state was deconvoluted, and raw data files were converted to .mgf by Proteome Discoverer version 1.4 (Thermo Scientific).

Data was processed using Mascot version 2.3.02 (Matrix Science) and searched against an in-house database (125561 sequences; 39895172 residues) comprising proteins *Panicum virgatum* v1.1 (DOE-JGI, http://phytozome.jgi.doe.gov) [8] and the common Repository of Adventitious Proteins (cRAP version 1.0, The Global Proteome Machine). Mascot was set to search with the following parameters: peptide tolerance of ± 10 ppm, MS/MS tolerance of ±

0.05 Da, variable modification with Oxidation (M), up to one missed cleavage for trypsin, and instrument type set to ESI-FTICR. Mascot search results were imported into Scaffold (v4.3.4, Proteome Software Inc.). Each biological replicate (comprising two independently analyzed technical replicates) was assigned as a BioSample in Scaffold using the parameters: LFDR scoring (all instruments) and protein cluster analysis. Matched data from each technical replicate was integrated by Scaffold to produce an integrated set of high confidence proteins identified in each biological replicate. Scaffold was used to further filter protein and peptide matches using the following parameters: peptide identifications greater than 95.0% probability by the Peptide Prophet algorithm [13] with Scaffold delta-mass correction, protein identifications greater than 99.0% probability, and protein identification containing at least 1 identified peptide.

A total of 2,514 unique switchgrass proteins were identified across all samples with a false discovery rate of 0.1 % (Table S1). This was comprised of 2,057 proteins identified from biological replicate Pv-Endo-1 and 2,207 proteins from biological replicate Pv-Endo-2 (Figure 1). There were 1,750 unique proteins identified in both replicates which comprise nearly 70 % of all identified proteins (Figure 1). Details of all matched peptides are outlined in Table S2. The RAW data associated with each biological and technical replicates have been deposited to the ProteomeXchange Consortium [14] via the PRoteomics IDEntifications database (PRIDE) partner repository with the dataset identifier PXD001351 and DOI 10.6019/PXD001351.

The 1,750 proteins identified in both biological replicates were profiled using Gene Ontology (GO) terms [15] which were obtained from Phytozome. Of the 1,750 identified proteins 1,436 (ca. 82 %) were associated with at least one GO annotation (Table S3). The GO terms from

the endomembrane proteins were profiled against GO terms from the entire switchgrass proteome (Figure 2). Proteins associated with organelles and membranes (Cellular Component) are significantly enriched in the proteome, as would be expected from such a preparation (Figure S1). The Cellular metabolic process (GO:0044237), Biosynthetic process (GO:0009058) and components associated with localization (Biological Process) are all significantly enriched; including, Establishment of localization (GO:0051234) with proteins associated with the level 5 child GO term, vesicle-mediated transport (GO:0016192) *e.g.* multiple RAB GTPase family members such as Pavir.Ca00563.1.p and Pavir.Ea02868.1.p. According to Molecular Function, Hydrolase activity (GO:0016787) is significantly over-represented which includes proteins in the level 4 child GO term, acting on glycosyl bonds (GO:0016798), *e.g.* glycoside hydrolases (GH) family members like, β-glucosidases (GH1: Pavir.J26569.1.p) and xyloglucan endotransglucosylase (GH16: Pavir.Ha01388.1.p). Such processes are typically associated with the plant endomembrane. Further switchgrass proteins likely to be associated with the plant endomembrane have been annotated (Table S3).

To further investigate the contribution this switchgrass endomembrane proteome dataset can provide, we selected candidates with a specific role in cell wall biosynthesis. Xylan is one of the most common polymers in the plant cell wall [16] and is synthesized in the Golgi apparatus from UDP-xylose. This activated donor is synthesized through decarboxylation of UDP-glucuronic acid (UDP-GlcA) by the enzyme UDP-Xylose Synthase (UXS) in both the cytosol and the Golgi lumen [17, 18]. We employed the UXS protein sequences from Arabidopsis and rice with BLAST to identify the UXS family in switchgrass. The switchgrass UXS family is comprised of 10 members, for which eight are putative homeologs (Figure S2). In total, the switchgrass endomembrane proteome identified seven UXS proteins between the two biological replicates. We were able to confidently discriminate between loci

for five members (Table 1). It is apparent from these data that rather than identifying a distinct homeolog (a paralog produced by polyploidy), it appears that both chromosomal versions are expressed (Table 1). If this observation generally holds for homeologs in switchgrass, it may have important implications for the genetic manipulation of switchgrass and polyploidy species in general. For example, since homeologs are likely to be functionally identical, all members will need to be genetically manipulated in order to produce true loss-of-function mutant line.

The development of the next generation biofuel crops will be crucial in supporting the burgeoning lignocellulosic biofuels industry. Underlying the improvement of these crops will be the creation of genetic resources and technologies to enable more sophisticated engineering and breeding strategies to be employed. This dataset highlights the contributions proteomics can provide to these resources as more plant genomes become publicly accessible. To the best of our knowledge, this dataset represents the first proteomic analysis of switchgrass material and showcases the maturity of the genomics data produced by next generation sequencing, even when used to address complex genomes such as switchgrass.

Acknowledgments

The work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy [DE-AC02-05CH11231]. JLH is supported by an Australian Research Council Future Fellowship [FT130101165]. RS and MKS acknowledge DBT, India for the Ramalingaswami Fellowship. BS is supported by a Human Frontier Science Program long-term post-doctoral fellowship [LT000674/2012]. We also wish to thank the UC Davis Proteomics Core Facility for sample analysis. Pre-publication data provided by the Department of Energy Joint Genome Institute.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Somerville, C., Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 2006, *22*, 53-78.

[2] Scheller, H. V., Jensen, J. K., Sorensen, S. O., Harholt, J., Geshi, N., Biosynthesis of pectin. *Physiol. Plantarum* 2007, *129*, 283-295.

[3] Scheller, H. V., Ulvskov, P., Hemicelluloses. *Annu. Rev. Plant Biol.* 2010, *61*, 263-289.

[4] Klein-Marcuschamer, D., Simmons, B. A., Blanch, H. W., Technoeconomic analysis of a lignocellulosic ethanol biorefinery with ionic liquid pretreatment. *Biofuels Bioprod. Biorefin.* 2011, *5*, 562-569.

[5] Ito, J., Petzold, C. J., Mukhopadhyay, A., Heazlewood, J. L., The role of proteomics in the development of cellulosic biofuels. *Curr. Proteomics* 2010, *7*, 121-134.

[6] Casler, M. D., Tobias, C. M., Kaeppler, S. M., Buell, C. R*., et al.*, The Switchgrass Genome: Tools and Strategies. *Plant Genome* 2011, *4*, 273-282.

[7] Sharma, M. K., Sharma, R., Cao, P. J., Jenkins, J*., et al.*, A Genome-Wide Survey of Switchgrass Genome Structure and Organization. *PLOS ONE* 2012, *7*.

[8] Goodstein, D. M., Shu, S., Howson, R., Neupane, R*., et al.*, Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012, *40*, D1178-1186.

[9] Hirsch-Hoffmann, M., Gruissem, W., Baerenfaller, K., pep2pro: the high-throughput proteomics data processing, analysis, and visualization tool. *Front. Plant Sci.* 2012, *3*, 123.

[10] Gibeaut, D. M., Carpita, N. C., Synthesis of (1→3), (1→4)-*β*-D-glucan in the Golgi apparatus of maize coleoptiles. *Proc. Natl Acad. Sci. USA* 1993, *90*, 3850-3854.

[11] Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 1976, *72*, 248-254.

[12] Parsons, H. T., Christiansen, K., Knierim, B., Carroll, A*., et al.*, Isolation and proteomic characterization of the Arabidopsis Golgi defines functional and novel targets involved in plant cell wall biosynthesis. *Plant Physiol.* 2012, *159*, 12-26.

[13] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002, *74*, 5383-5392.

[14] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A*., et al.*, ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 2014, *32*, 223-226.

[15] Gene Ontology Consortium, Gene Ontology: tool for the unification of biology. *Nat. Genet.* 2000, *25*, 25-29.
[16] Rennie, E. A., Scheller, H. V., Xylan biosynthesis. *Curr. Opin. Biotech.* 2014, *26*, 100-107.
[17] Harper, A. D., Bar-Peled, M., Biosynthesis of UDP-xylose. Cloning and characterization of a novel Arabidopsis gene family, UXS, encoding soluble and putative membrane-bound UDP-glucuronic acid decarboxylase isoforms. *Plant Physiol.* 2002, *130*, 2188-2198.
[18] Pattathil, S., Harper, A. D., Bar-Peled, M., Biosynthesis of UDP-xylose: characterization of membrane-bound *AtUxs2*. *Planta* 2005, *221*, 538-548.
[19] Ye, J., Fang, L., Zheng, H. K., Zhang, Y.*, et al.*, WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 2006, *34*, W293-W297.
[20] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H.*, et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997, *25*, 3389-3402.
[21] Suzuki, K., Watanabe, K., Masumura, T., Kitamura, S., Characterization of soluble and putative membrane-bound UDP-glucuronic acid decarboxylase (OsUXS) isoforms in rice. *Arch. Biochem. Biophys.* 2004, *431*, 169-177.
[22] Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D.*, et al.*, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 2012, *40*, D1202-1210.
[23] Ouyang, S., Zhu, W., Hamilton, J., Lin, H.*, et al.*, The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 2007, *35*, D883-887.
[24] Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., Millar, A. H., SUBA: the Arabidopsis Subcellular Database. *Nucleic Acids Res.* 2007, *35*, D213-218.
[25] Dereeper, A., Guignon, V., Blanc, G., Audic, S.*, et al.*, Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008, *36*, W465-469.

Figure Legends

Figure 1. Experimental workflow for the proteomic characterization of the endomembrane from switchgrass coleoptiles.

The proteins identified at each step correspond to numbers after data filtering by Scaffold.

Figure 2. Gene Ontology terms assigned to the 1750 switchgrass endomembrane proteins.

The top ten GO annotations (level 3) assigned in each of cellular component, Biological Process and Molecular Function categories with p value ≤ 0.001 (Pearson Chi-Square test) are shown in the form of a histogram. Analysis was conducted using the WEGO tool [19] using default parameters.

Table 1. The UDP-xylose synthase (UXS) family from switchgrass.

Putative switchgrass UXS homeologs are shown as color coded pairs, unassigned members are not highlighted. The homeologs and percent identity was determined by reciprocal BLAST [20]. The UXS classifications were obtained from [21] after BLAST analysis against *Arabidopsis thaliana* release TAIR10 [22] and *Oryza sativa* (rice) release v 7.0 [23]. The subcellular location was inferred based on experimental information for Arabidopsis homologs [24]. The unique peptide match for each locus is outlined in Table S4.

Switchgrass locus[a]	Chromosome number	Top BLAST match	% identity with putative homeolog	UXS Classification	Inferred subcellular location	Unique Peptides	
						Pv-Endo-1	Pv-Endo-2
Pavir.Ib01205.1.p	9b	Pavir.Ia03662.1	97	Type UXS3	cytosol	2	2
Pavir.Ia03662.1.p	9a	Pavir.Ib01205.1	97	Type UXS3	cytosol	2	2
Pavir.J11244.1.p	contig13802	Pavir.Cb01376.1	97	Type UXS2	Golgi	-	1
Pavir.Cb01376.1.p	3b	Pavir.J11244.1	98	Type UXS2	Golgi	-	1
Pavir.Ea03089.1.p[b]	5a	Pavir.Eb03654.1.p	99	Type UXS1	Golgi	-	2
Pavir.Eb03654.1.p[b]	5b	Pavir.Ea03089.1.p	99	Type UXS1	Golgi	-	2
Pavir.J34746.1.p	contig48762	Pavir.Ib01306.1	98	Type UXS1	Golgi	-	-
Pavir.Ib01306.1.p	9b	Pavir.J34746.1.p	98	Type UXS1	Golgi	-	-
Pavir.Ba00210.1.p	2a	Pavir.Bb03771.1	97	Type UXS2	Golgi	-	-
Pavir.Eb02349.1.p	5b	Pavir.J11242.1	85	Type UXS2	Golgi	2	5

[a]Locus code corresponds to *Panicum virgatum* assembly v1.1 from Phytozome [8].

[b]These switchgrass homeologs were identified by identical peptides and were indistinguishable.

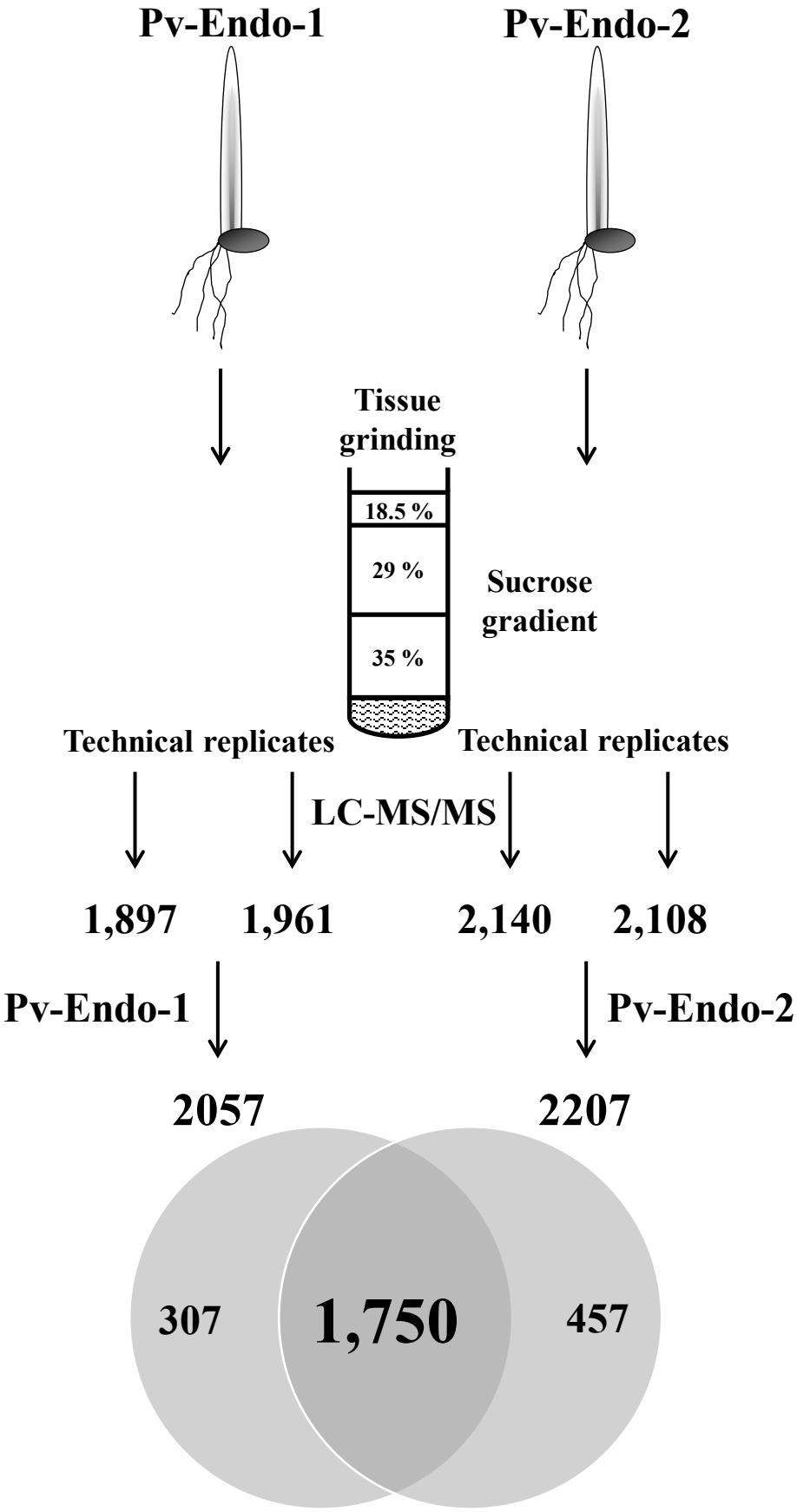
Pv-Endo-1
Pv-Endo-2
Tissue grinding
18.5 %
29 %
35 %
Sucrose gradient
Technical replicates
Technical replicates
LC-MS/MS
1,897
1,961
2,140
2,108
Pv-Endo-1
Pv-Endo-2
2057
2207
307
1,750
457

Fig 1

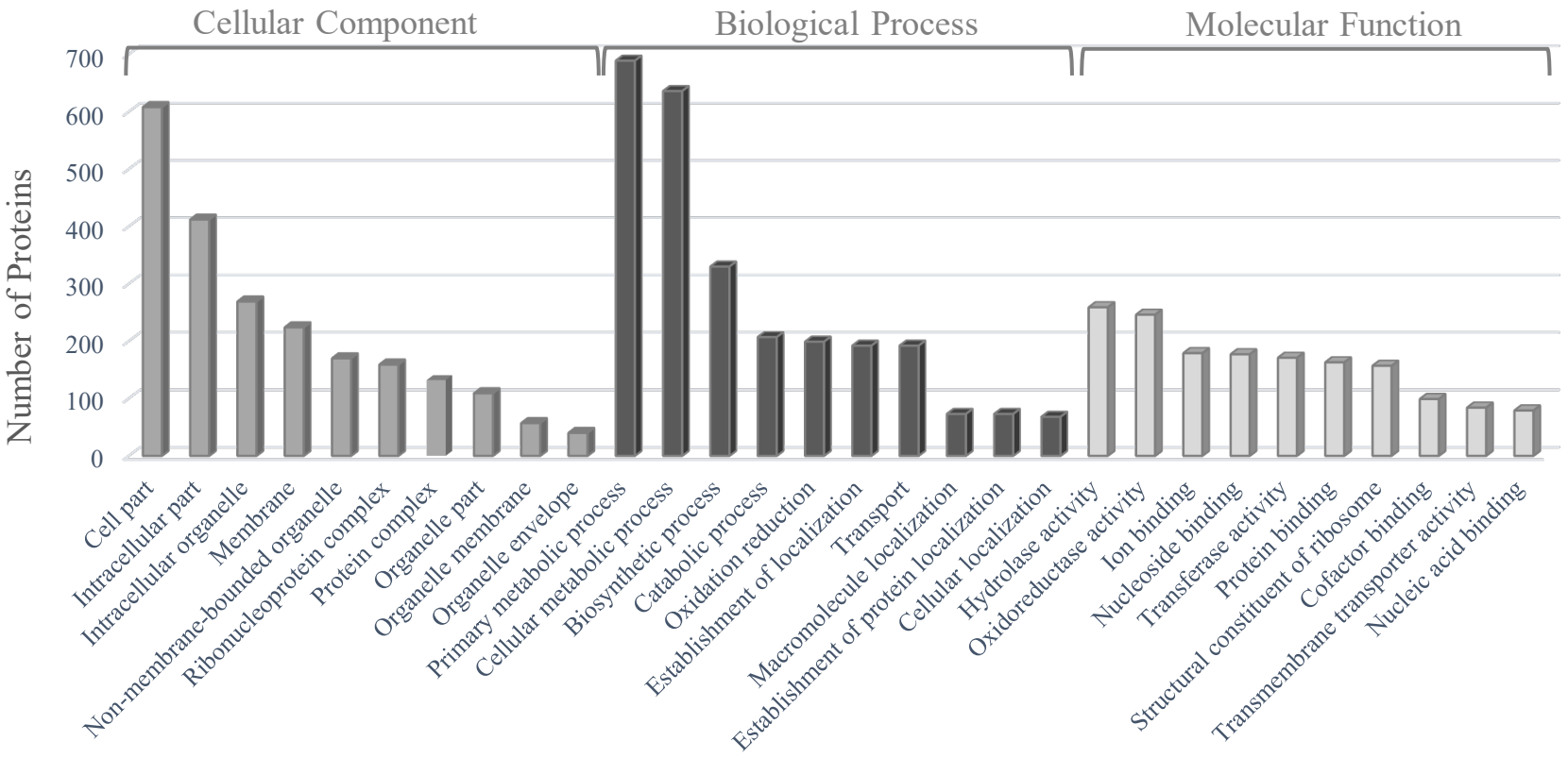
Cellular Component
Biological Process
Molecular Function
Number of Proteins
700
600
500
400
300
200
100
0
Cell part
Intracellular part
Intracellular organelle
Membrane
Non-membrane-bounded organelle
Ribonucleoprotein complex
Protein complex
Organelle part
Organelle membrane
Organelle envelope
Primary metabolic process
Cellular metabolic process
Biosynthetic process
Catabolic process
Oxidation reduction
Establishment of localization
Transport
Macromolecule localization
Establishment of protein localization
Cellular localization
Hydrolase activity
Oxidoreductase activity
Ion binding
Nucleoside binding
Transferase activity
Protein binding
Structural constituent of ribosome
Cofactor binding
Transmembrane transporter activity
Nucleic acid binding

Fig. 2

Author/s:
Lao, J; Sharma, MK; Sharma, R; Fernandez-Nino, SMG; Schmutz, J; Ronald, PC; Heazlewood, JL; Schwessinger, B

Title:
Proteome profile of the endomembrane of developing coleoptiles from switchgrass (Panicum virgatum)

Date:
2015-07-01

Citation:
Lao, J; Sharma, MK; Sharma, R; Fernandez-Nino, SMG; Schmutz, J; Ronald, PC; Heazlewood, JL; Schwessinger, B, Proteome profile of the endomembrane of developing coleoptiles from switchgrass (Panicum virgatum), PROTEOMICS, 2015, 15 (13), pp. 2286 - 2290