

PPDB, the Plant Proteomics Database at Cornell

Qi Sun¹, Boris Zybailov², Wojciech Majeran², Giulia Friso²,
Paul Dominic B. Olinares² and Klaas J. van Wijk^{2,*}

¹Computation Biology Service Unit, Cornell Theory Center and ²Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA

Received August 15, 2008; Revised September 17, 2008; Accepted September 18, 2008

ABSTRACT

The Plant Proteomics Database (PPDB; <http://ppdb.tc.cornell.edu>), launched in 2004, provides an integrated resource for experimentally identified proteins in *Arabidopsis* and maize (*Zea mays*). Internal BLAST alignments link maize and *Arabidopsis* information. Experimental identification is based on in-house mass spectrometry (MS) of cell type-specific proteomes (maize), or specific subcellular proteomes (e.g. chloroplasts, thylakoids, nucleoids) and total leaf proteome samples (maize and *Arabidopsis*). So far more than 5000 accessions both in maize and *Arabidopsis* have been identified. In addition, more than 80 published *Arabidopsis* proteome datasets from subcellular compartments or organs are stored in PPDB and linked to each locus. Using MS-derived information and literature, more than 1500 *Arabidopsis* proteins have a manually assigned subcellular location, with a strong emphasis on plastid proteins. Additional new features of PPDB include searchable posttranslational modifications and searchable experimental proteotypic peptides and spectral count information for each identified accession based on in-house experiments. Various search methods are provided to extract more than 40 data types for each accession and to extract accessions for different functional categories or curated subcellular localizations. Protein report pages for each accession provide comprehensive overviews, including predicted protein properties, with hyperlinks to the most relevant databases.

INTRODUCTION

The field of plant proteomics has greatly accelerated over the last years, in particular due to advances in mass spectrometry (MS)-based techniques, as well as associated bioinformatics tools to search the experimental MS data

against the various plant genomes (e.g. *Arabidopsis* and rice) and EST assemblies (e.g. maize, tomato) (1,2). Currently, there are at least 80 small- to large-scale MS-based experimental proteomics datasets published for *Arabidopsis* and multiple studies for other plant species, in particular maize, rice and *Medicago truncatula* (see further below). In addition, there are many original publications on the function and (subcellular) localization of plant proteins, with the majority concerning *Arabidopsis*. These subcellular localizations are determined by immuno-detection techniques, protein import assays or through visualization of GFP/YFP fusion proteins. In order to assemble and comprehend all this protein information, various plant proteomics databases have been developed, with each having certain strengths or emphasizing particular plant species or (sub)cellular compartments. One of the major challenges is to extract experimental protein information from the literature or from local data repositories and use this information to annotate protein function and subcellular localization.

The Plant Proteomics Database, PPDB, was launched in 2004. Initially PPDB was named Plastid Proteome DB as it was dedicated to plant plastids with the objective of disseminating our chloroplast proteomics data and integrating these with other types of proteomics information (3). Since its inception in 2004, the PPDB interface and its content have greatly expanded. To better reflect this expansion, we have recently renamed the database as Plant Proteome DB even if most efforts regarding manual curation (name, function and localization) are still focused on the plastid. PPDB is a unique resource for the plant community and complements other plant proteomics resources, such as the plant mass spectral reference database Promex (4), SUBA for *Arabidopsis* protein localization (5), as well as databases specialized in other organelles, such as peroxisomes (6) or phosphopeptides (7). Direct links at the protein locus level to the most relevant databases are present in PPDB.

The central objective of the PPDB is to provide in-house experimental MS-based information for cell type-specific or subcellular proteomes in maize and *Arabidopsis*, as well as their predicted properties, and integrate this with

*To whom correspondence should be addressed. Tel: +1 607 255 3664; Fax: +1 607 255 5407; Email: kv35@cornell.edu

published experimental data from other (external) sources. Importantly, information from in-house MS-based identifications and posttranslational modifications (PTMs) is available for each identified protein accession. (Note: throughout this article and in the PPDB, we use the term ‘accession’ to describe the identifier of genes and proteins.) This allows the database user to determine the significance of the experimental identifications and also evaluate information regarding PTMs. Multiple search methods are provided so that the user can retrieve information based on accession or protein name, functional annotation or various protein properties or experiments. The annotation for each accession is enhanced by manual curation. Below, we review the internal and external data (sources) in some detail (see also Figures 1–3 and Supplementary Figure 1).

DATABASE STRUCTURE, DATA SOURCES AND CURATION

The database engine is a Microsoft SQL server. The web interface was developed on ASP.NET platform using C# language. All protein-encoding gene models in the *Arabidopsis* nuclear and organellar genomes as assembled by TAIR (<http://www.arabidopsis.org/>) (currently release Ath8.0), as well as all maize EST assemblies (ZmGI) by TIGR (<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=maize>) and the draft maize genome (<http://maizesequence.org/>). These are all uploaded in PPDB and are linked to each other via a BLAST alignment (Figure 1). Thus every predicted protein in both species can be searched for experimental and other relevant information, even if not experimentally identified.

In-house MS and proteomics datasets

All in-house experimental data are from different compartments from within the chloroplast, such as thylakoids, plastoglobules, stroma, as well as whole leaves or specific cell types or structures [e.g. bundle sheath (BS) strands] from *Arabidopsis* and maize (Figure 1). In the case of maize, a C4 plant, chloroplasts are isolated from either BS cells or mesophyll (M) cells and their proteomes analyzed to address cell-type-specific specialization (Figure 1). Basic information about each experiment can be found in experimental descriptions, including the publication where the data were presented if applicable. Currently there are more than 140 available in-house experiments. Most of these experiments involve gel-based protein separations (native or denaturing) and in-gel digestions with trypsin, while a smaller set of experiments is based on in-solution tryptic digestion.

The vast majority of in-house MS-based identifications are from LC-ESI-MS/MS, either using a Q-TOF (Waters, Milford, MA) or an LTQ-Orbitrap (ThermoFisher, Madison, WI), with the remainder from peptide mass finger printing using a MALDI-TOF MS instrument (Applied Biosystems, Foster City, CA) (Figure 1). The instrument settings and database searches (always using Mascot) and filter criteria (aimed at <1% false positive rate for peptide identification, with cutoff set by decoy database searches (8)) are standardized (9–11). A large number of output parameters from the Mascot searches [accession number, experimental sequence ambiguity, Mowse score, number of matching peptides, number of matched MS/MS spectra (queries), number of unique queries, highest peptide score, lowest precursor peptide error (in p.p.m.), sequence coverage, tryptic status and determined peptide modification states] can be retrieved

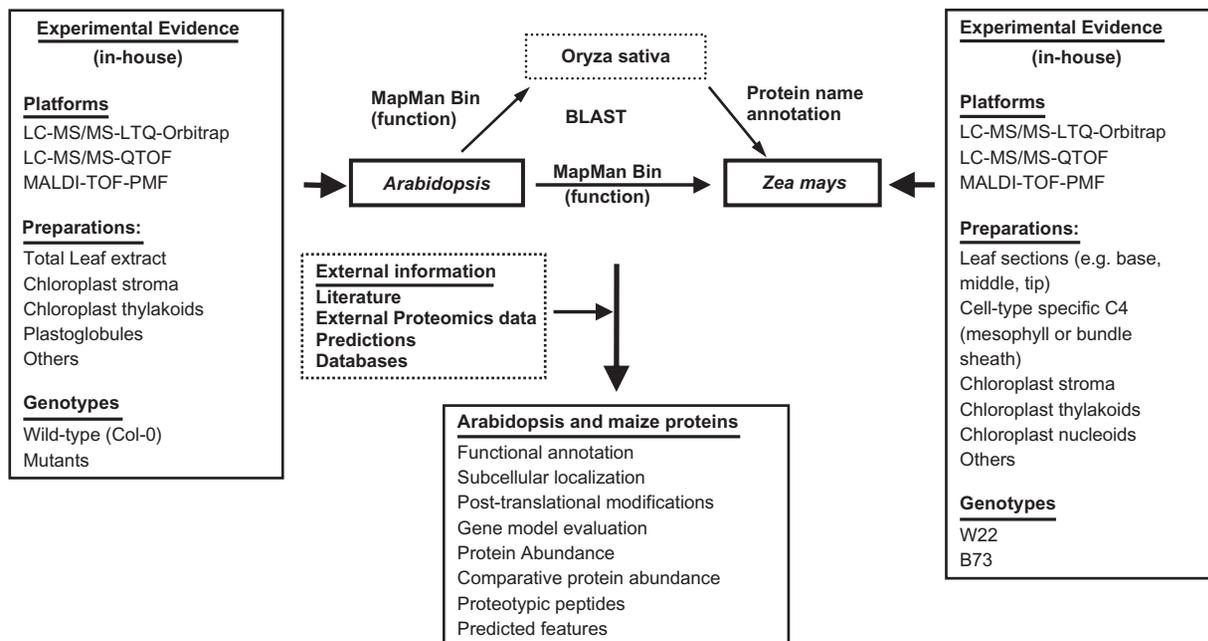


Figure 1. Basic flow diagram of the information in PPDB.

from the PPDB (see further below). For all peak lists obtained by LTQ-Orbitrap, three parallel searches are performed: (i) *tryptic* search with precursor ion tolerance window set at ± 6 p.p.m., (ii) *error-tolerant* search with precursor ion tolerance window set at ± 3 p.p.m., (iii) *semi-tryptic* search with acetylation of peptide N-terminus set as a variable modification (9). After these searches, all redundant queries are removed and only the query with the highest ion score for a precursor ion with MS/MS ions are kept and uploaded into PPDB. A manuscript with more details about this procedure and the significance of the various PTMs that can be detected without specific enrichment procedures is in preparation (B.Z., Q.S. and K.J.vW., unpublished data). In-house MS data from previous publications (since summer 2004) were all re-searched and filtered using the same standards, and search results were uploaded into PPDB.

External (published) proteomics datasets. Currently more than 80 small- to large-scale *Arabidopsis* (and a few other *Brassicacea*) proteome datasets are stored in PPDB and linked to each locus. These data sets originate from various subcellular compartments (e.g. plasma membrane, vacuole, chloroplasts) and organs (e.g. leaf, root) or cell types (e.g. suspension cells, epidermis, trichome). This information can be obtained by selecting 'Proteomics Publication' as output parameter. The complete list of published proteomics papers and each of their accessions can be downloaded from PPDB. These data are also displayed on the relevant 'protein report' pages (see below).

Subcellular localization, functional annotation and manual curation. To determine the subcellular localization of proteins, we cross-correlated our in-house MS-based identifications to more than 80 published proteomics papers on *Arabidopsis* subcellular fractions, as well as information extracted from TAIR and published studies providing details about experimental protein localization. As a rule of thumb, subcellular localization by GFP/YFP and western blots was considered strong evidence, although we noted that there are several examples of incorrect subcellular localization assignment based on GFP/YFP. Identification in published proteomics studies was sometimes difficult to judge since information about the confidence of MS-based identification was not easily accessible. Subcellular localization is not assigned if there is insufficient or too much conflicting evidence.

Each protein is assigned a molecular function, using the hierarchical, non-redundant classification system developed for MapMan (12). Where possible, functional assignments are verified manually and additional new functional categories (Bins) are created, if needed. Since many of the maize ZmGI accessions and all loci on the maize genome lack functional annotation, we functionally annotated all identified ZmGI accessions using a combination of best BLAST hits in the predicted rice proteome (OSGI), the predicted *A. thaliana* proteome, ATHv8, further supported by BLAST searches against the nonredundant NCBI database. We are currently in the process of providing a tentative gene name annotation and function (MapMan bin) for all predicted proteins in the maize

genome draft; this will be based on best BLAST hits to the rice gene index (OsGI v4) for the protein name and best BLAST hit to the *Arabidopsis* proteome (Ath v8) for assignment of the MapMan bins (Figure 1).

In-house predicted protein properties and subcellular localization

Predicted chloroplast localization and predicted chloroplast transit peptide (cTP) and luminal transit peptide (LTP), as well as various predicted physical-chemical properties (e.g. pI, mass, hydrophobicity, trans-membrane domains, cystein content, etc.) of precursor and processed proteins are provided for each *Arabidopsis* accession. Details for these predictions can be found in Refs (3,13).

PPDB SEARCH FUNCTIONS AND EXTRACTABLE PROTEIN INFORMATION

The PPDB has nine search functions to extract multiple types of information (output) stored in PPDB for any accession. Extraction of the desired output can be performed by simply choosing a search function (Figure 2), and selecting the relevant 'check boxes' (Figure 3, Supplementary Material). Information can be extracted from the PPDB for individual accessions or in batch format. An overview of the major search functions and the input is shown in Figure 2. Each search can be restricted to a specific experiment, or groups of experiments, and also to a specific species (maize or *Arabidopsis*) or the source of the accession. More than 40 data types can be selected as shown in Figure 3. The output of each search is a list of accessions with the data types that were selected. Accessions are hyperlinked to their respective protein report pages.

PROTEIN-REPORT PAGES—A CENTRAL TOOL OF THE PPDB

Information for each protein in PPDB is summarized in a 'protein-report page', thus providing an integrated overview of key information for each protein (Supplementary Figure 1A–D). The page summarizes information about (curated) subcellular localization, function, homologs, predicted protein properties, in-house experimental MS-based identifications and cross-references to published studies in which the protein was identified. Relevant databases (for *Arabidopsis*: TAIR, AtProteome, *PhosPhAt*, ProMex, SUBA and POGS/PlantRBP; for maize: maizesequence.org) are hyperlinked for each accession to rapidly obtain additional information. These report pages also provide detailed information about matched MS data and how it maps to the protein. For individual protein searches, this is the best way to obtain a comprehensive overview. Supplementary Figure 1 shows an example of a protein-report page for histidinol dehydrogenase (At5g63890) localized in the chloroplast stroma.

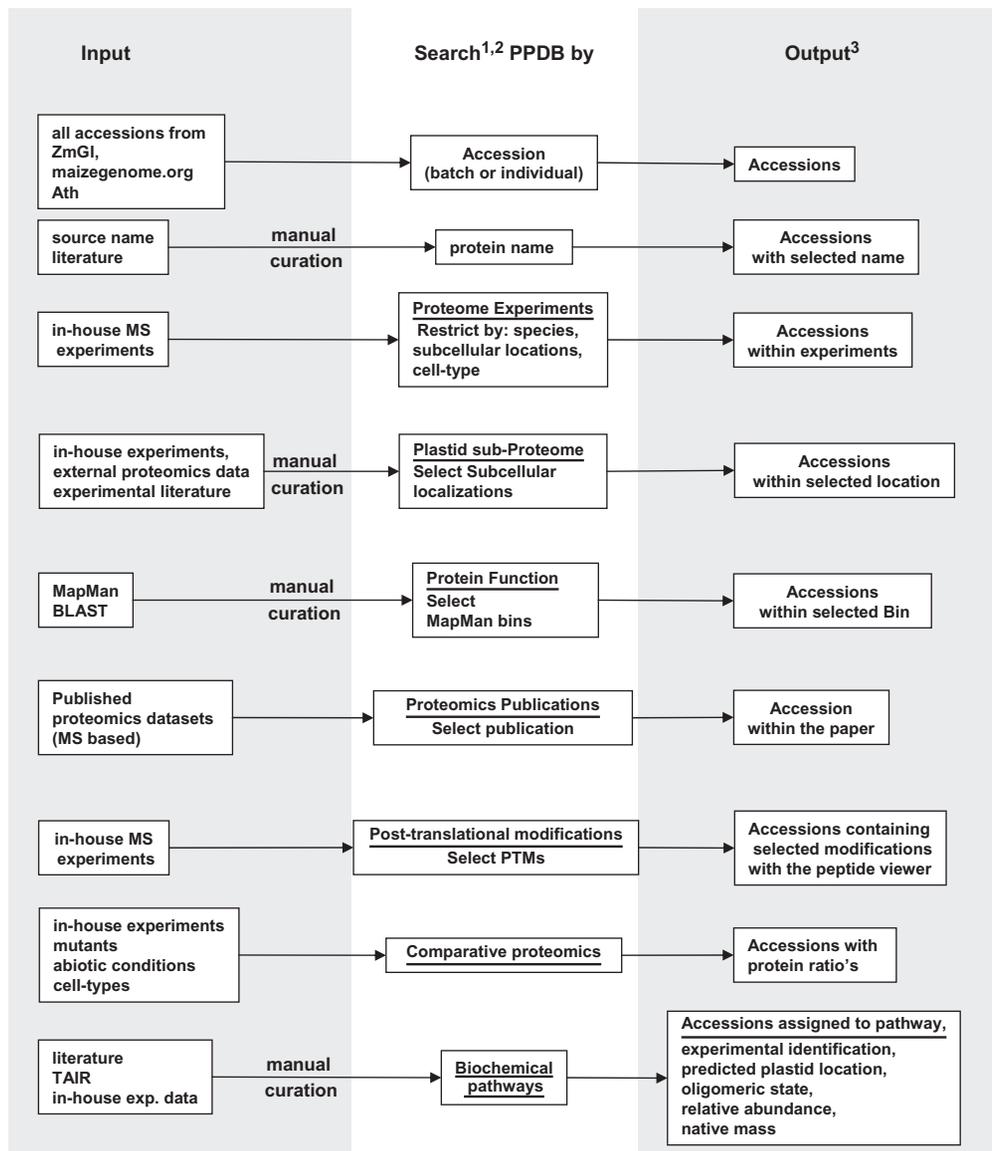


Figure 2. Input, output and the main search functions in PPDB. There are nine main search functions in PPDB each facilitating the search of specific sets of information and providing different ways to interrogate the database. The input for each of these search functions is listed. ¹Searches can be restricted to specific experiment, species or source of accession. ²More than 40 data types can be selected. ³More than 40 data types for each accession can be displayed. Accessions are hyperlinked to their respective protein-report page.

PROJECTED MS DATA ON PREDICTED PROTEIN MODELS

Peptides identified by MS are projected onto predicted protein models using 'pop-up' windows available at each protein-report page (Supplementary Figure 1B–D). This allows the user to better evaluate the significance of the protein identification and the relevance of multiple gene models (if present; e.g. see Supplementary Figure 1B) and PTMs. Predicted cTPs are indicated in the protein models and comparison between the most N-terminal identified (tryptic or semi-tryptic) peptide and the predicted cTP will aid in understanding the subcellular localization (Supplementary Figure 1C and D). The complete list of all identified peptide sequences with possible PTMs, with

the experimental identifier, is listed in the same 'pop-up' window for each identified protein in both maize and *Arabidopsis* (Supplementary Figure 1C). A list of peptides identified using the error-tolerant search (in Mascot) is provided as an additional option. An overview of the identified peptides can also be restricted to a selected experiment (in this case experiment #451) (Supplementary Figure 1D). This window provides experimental details, such as the type of mass spectrometer used for protein identification and for each identified peptide the charge state, mass error, ion score, the number of time this peptide was identified in this sample, precursor ion intensity and type of search (full tryptic, semi-tryptic or error-tolerant search). Graphic displays that map the identified

Display options: ?

- | | | | | |
|---|---|---|---|---|
| <input type="checkbox"/> STD. Annot. | <input type="checkbox"/> Lab Annot. | <input type="checkbox"/> Curated Loc. | <input type="checkbox"/> TAIR Loc. | <input type="checkbox"/> MAPMANBin |
| <input type="checkbox"/> Sep. MAPMANBin | <input type="checkbox"/> Aramemnon | <input type="checkbox"/> TMHMM | <input type="checkbox"/> TargetP | <input type="checkbox"/> Predotar |
| <input type="checkbox"/> PFAM | <input type="checkbox"/> Cys | <input type="checkbox"/> Calc.gravy | <input type="checkbox"/> Calc.MW | <input type="checkbox"/> Calc.PI |
| <input type="checkbox"/> exp.MW_d | <input type="checkbox"/> exp.pI_d | <input type="checkbox"/> exp.MW_n | <input type="checkbox"/> exp.pI_n | <input type="checkbox"/> exp.Volume |
| <input type="checkbox"/> exp.ambiguity | <input type="checkbox"/> exp.amb_gmodel | <input type="checkbox"/> exp.MS_type | <input type="checkbox"/> exp.MascotScore | <input type="checkbox"/> exp.#peptides |
| <input type="checkbox"/> exp.ID* | <input type="checkbox"/> exp.spot | <input type="checkbox"/> exp.queries | <input type="checkbox"/> exp.uniq.queries | <input type="checkbox"/> exp.Tissue |
| <input type="checkbox"/> exp.Sep.Steps | <input type="checkbox"/> CellType | <input type="checkbox"/> ProteomicsPub. | <input type="checkbox"/> GFP/YFP loc. | <input type="checkbox"/> Best Rice hit |
| <input type="checkbox"/> Best Maize hit | <input type="checkbox"/> Coverage | <input type="checkbox"/> Top3pep | <input type="checkbox"/> Top3pep(-MetCys) | <input type="checkbox"/> #Tryptic Frag. |
| <input type="checkbox"/> #Tryptic Frag.(-ctp) | <input type="checkbox"/> #Tryptic Frag.(-ltp) | | | |

*exp.ID: [Details of the experimental sources.](#)

Output:

- web excel file

Restrict species:

Species: v

Source of sequence: v

Extra options:

Experiments Spots

(enter experiment IDs separated by space. [Check the list](#) if you do not know the IDs.)

Figure 3. Output options for the various search functions can be selected via check boxes. The output can be provided as a webpage or as an Excel file. The various search functions can be restricted by plant species (currently *Arabidopsis* and maize) and also by the source of the sequence database (e.g. Ath, Tigr, maizegenome.org). Additionally, searches can be restricted to particular (sets of) experiments. See further the Supplementary Material for more explanations.

peptides onto the predicted exons for all available protein models are provided as 'pop-up' windows on each protein-report page (Supplementary Figure 1B).

PROTEIN ABUNDANCE BY SPECTRAL COUNTS AND PROTEOTYPIC PEPTIDES

Recently, large-scale MS-based studies for yeast, humans, *Escherichia coli* and other sequenced organisms have shown that the number of MS/MS spectra matched to a protein (spectral counts) positively correlates with the protein abundance (14–17). Upon control of several experimental conditions, careful and stringent spectral assignments, and sophisticated normalization procedures, it appears that MS-based quantification can provide an attractive and sensitive tool to obtain large-scale measurements of relative protein concentrations. For further review and discussions we refer to Refs (18–20). In a recent paper, we showed that 'spectral counting' can indeed provide large-scale protein quantification for *Arabidopsis* (9), if experiments are carefully designed with attention to reproducibility at every step of the process and if appropriate thresholds are applied for the minimum number of matched spectral counts for each accession and low false positive peptide identification rates. Moreover, normalization and removal of redundant queries (also named spectral counts) and corrections for shared peptides are also important for accurate quantification.

To obtain a qualitative view of protein abundance in the various protein preparations, the number of spectral counts (or queries) for each identified accession can be extracted by selection of the relevant 'check boxes' in the output menu [experimental queries and unique queries (Figure 3)]. In addition, these peptide sequences and frequency can also be found displayed graphically via 'pop-up' windows at each protein-report page. The most frequently observed peptides for a protein accession can serve as a (quantitative) signature, if this peptide uniquely matched to the accession (proteotypic peptide). The most frequently observed peptides (the top three) can be extracted for each accession by selecting it as an output parameter (Top3pep). Moreover, if such peptides are to be used for quantification, the selection may be further constrained to peptides that do not contain cysteine nor methionine residues, since they are prone to modifications, leading to unreliable quantification (Top3pep-MetCys) (Figure 3).

FUTURE DIRECTIONS

The PPDB is continuously updated with new in-house experiments, as well as external data sets. Manual assignment of function and subcellular localization for proteins identified by in-house experiments or new proteins discovered in plastids is also regularly performed (at least monthly). When the sequencing and assembly of the

maize genome is in a more advanced state, we will move from using the ZmGi as database for searching the MS data for maize samples to the maize genome sequence. Currently, we are searching both sets of maize sequences in parallel. We will therefore make an effort to incorporate protein names for these new maize protein accessions, and assign functions (MapMan bins) as well as subcellular localization. Several in-house experiments involve a quantitative comparative analysis of cell-type-specific differences in maize (11,21) and differences between chloroplast mutants and wild-type plants in *Arabidopsis* (10,22). Some of this information is displayed in PPDB either in the relevant protein-report pages or per experiment. Work is in progress to improve this function.

Finally, we aim to keep working closely with other plant community databases (e.g. TAIR, Gramene and others) and colleagues around the world to distribute our data and provide efficient links.

AVAILABILITY

PPDB can be accessed at <http://ppdb.tc.cornell.edu/>. The software for the PPDB database and web site is available upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The PPDB and other bioinformatics infrastructure were generated using the resources of the Cornell Theory Center. We thank the curators of MapMan bin for providing us all MapMan bin assignments to *Arabidopsis* accessions.

FUNDING

The National Science Foundation (MCB 0343444; MCB-0718897; PGRP-0211935); the US Department of Energy (DE-FG02-04ER15560); the New York State Office of Science, Technology and Research (NYSTAR to K.J.V.W.); Cornell University, New York State, federal agencies, foundations and corporate partners. Funding for open access charge: National Science Foundation grant PGRP-0701736.

Conflict of interest statement. None declared.

REFERENCES

- Rosignol, M., Peltier, J.B., Mock, H.P., Matros, A., Maldonado, A.M. and Jorin, J.V. (2006) Plant proteome analysis: a 2004–2006 update. *Proteomics*, **6**, 5529–5548.
- Jorin, J.V., Maldonado, A.M. and Castillejo, M.A. (2007) Plant proteome analysis: a 2006 update. *Proteomics*, **7**, 2947–2962.
- Friso, G., Giacomelli, L., Ytterberg, A.J., Peltier, J.B., Rudella, A., Sun, Q. and Wijk, K.J. (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell*, **16**, 478–499.
- Hummel, J., Niemann, M., Wienkoop, S., Schulze, W., Steinhauser, D., Selbig, J., Walther, D. and Weckwerth, W. (2007) ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics*, **8**, 216.
- Heazlewood, J.L., Tonti-Filippini, J., Verboom, R.E. and Millar, A.H. (2005) Combining experimental and predicted datasets for determination of the subcellular location of proteins in *Arabidopsis*. *Plant Physiol.*, **139**, 598–609.
- Reumann, S., Ma, C., Lemke, S. and Babujee, L. (2004) AraPeroX. A database of putative *Arabidopsis* proteins from plant peroxisomes. *Plant Physiol.*, **136**, 2587–2608.
- Heazlewood, J.L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D. and Schulze, W.X. (2008) PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res.*, **36**, D1015–1021.
- Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. and Gygi, S.P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.*, **2**, 43–50.
- Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q. and van Wijk, K.J. (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE*, **3**, e1994.
- Rutschow, H., Ytterberg, A.J., Friso, G., Nilsson, R. and van Wijk, K.J. (2008) Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in *Arabidopsis thaliana*. *Plant Physiol.*, **148**, 156–175.
- Majeran, W., Zybailov, B., Ytterberg, A.J., Dunsmore, J., Sun, Q. and van Wijk, K.J. (2008) Consequences of C4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells. *Mol. Cell. Proteomics*, **7**, 1609–38.
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y. and Stitt, M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.*, **37**, 914–939.
- Sun, Q., Emanuelsson, O. and van Wijk, K.J. (2004) Analysis of curated and predicted plastid subproteomes of *Arabidopsis*. Subcellular compartmentalization leads to distinctive proteome properties. *Plant Physiol.*, **135**, 723–734.
- Liu, H., Sadygov, R.G. and Yates, J.R. 3rd. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.*, **76**, 4193–4201.
- Zybailov, B., Coleman, M.K., Florens, L. and Washburn, M.P. (2005) Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.*, **77**, 6218–6224.
- Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A. and Ahn, N.G. (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics*, **4**, 1487–1502.
- Lu, P., Vogel, C., Wang, R., Yao, X. and Marcotte, E.M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.*, **25**, 117–124.
- Listgarten, J. and Emili, A. (2005) Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics*, **4**, 419–434.
- Nesvizhskii, A.I., Vitek, O. and Aebersold, R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods*, **4**, 787–797.
- Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.*, **389**, 1017–1031.
- Majeran, W., Cai, Y., Sun, Q. and van Wijk, K.J. (2005) Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. *Plant Cell*, **17**, 3111–3140.
- Giacomelli, L., Rudella, A. and van Wijk, K.J. (2006) High light response of the Thylakoid proteome in *Arabidopsis* wild type and the ascorbate-deficient mutant *vtc2-2*. A comparative proteomics study. *Plant Physiol.*, **141**, 685–701.