2008 Proteomics Subcommittee Report

Prepared by Wolfram Weckwerth (Co-chair, weckwerth@mpimpgolm.mpg.de), Sacha Baginsky (Co-chair, sbaginsky@ethz.ch), Klaas van Wijk (kv35@cornell.edu) and Harvey Millar (Co-chair, harvey.millar@uwa.edu.au)

In 2006 a MASC subcommittee for *Arabidopsis thaliana* proteomics was established to consolidate databases, technique standards and experimentally validated candidate genes and functions. Since that time many new approaches and databases were developed. Altogether the resources of the MASC Proteomics subcommittee provide the largest collection of proteomics data for this higher model plant. Below we have listed the achievements of the last two years.

- A common webpage "MASC proteomics" was established including standards for different proteomic techniques, databases, procedures, meetings, proteome labs, etc. (see http://www.masc-proteomics.org/). This webpage will also host discussion platforms and database-crosslinks in the future.

- The largest collection of proteomics data for a single study in *Arabidopsis thaliana* was published in Science, 2008 (Baginsky lab), and assembled as an accessible database (www.AtProteome.ethz.ch). Links to TAIR (www.arabidopsis.org/), MIPS (http://mips.gsf.de) and TIGR (http://www.tigr.org/) were established. The data were further deposited in PRIDE and PRIDE Biomart (see http://www.ebi.ac.uk/pride/). The Arabidopsis proteome map provides expression evidence for close to 50% of all predicted gene models together with a number of alternative gene models. A set of organ specific biomarkers is provided together with organ-specific proteotypic peptides for 4,105 proteins to facilitate targeted quantitative proteomics surveys. Quantitative information for the identified proteins established correlations between transcript and protein accumulation in different plant organs. Further plans are to link the AtProteome database to existing resources in the MASC proteomics committee, such as PPDB (see http://ppdb.tc.cornell.edu/) and ProMEX (see http://promex.mpimp-golm.mpg.de/home.shtml) (see below).

- The most comprehensive study on the Arabidopsis chloroplast proteome, their sorting signals and PTMs as well as protein abundance, using high accuracy mass spectrometry (Orbitrap), was published 2008 in PLOS-One (van Wijk lab). These and other data are available via the plant proteome database (PPDB – link above) for Arabidopsis and maize. PPDB provides genome-wide experimental and functional characterization of the proteomes, including PTMs and subcellular localization, with emphasis on leaf and plastid proteins. Maize and Arabidopsis proteome information are directly linked via internal BLAST alignments and each protein is linked to TAIR, SUBA, ProMEX.

- The most comprehensive database on subcellular localization of proteins–SUBA- was established by the Millar lab (www.plantenergy.uwa.edu.au/applications/suba/index.php). These data are cross-linked to TAIR protein pages (www.arabidopsis .org) and selected data are provided as web services via the BioMoby Dashboard.

- The most comprehensive database of published and predicted phosphorylation sites in Arabidopsis was established, called PhosPhAT (see http://phosphat.mpimp-golm.mpg.de/) (Heazlewood/Weckwerth/Schulze lab). This database is crosslinked to ProMEX.

- ProMEX: a central searchable database of MS/MS reference spectra derived from *Arabidopsis thaliana, Chlamydomonas reinhardtii, Medicago truncatula,* potato, tomato and other proteomics samples was established and cross-referenced to the UNIPROT plant genome annotation initiative in 2007 (see http://promex.mpimp-golm.mpg.de/home.shtml) (Weckwerth lab). This database also facilitates the design of proteotypic peptides for targeted accurate protein quantification in complex samples. ProMEX has also crosslinks to PhosPhAT, PPDB and the Golmer Metabolome database (see GMD http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html). Links to AtProteome are in progress.

- AnnoJ: a Web 2.0 genome browser developed by Julian Tonti-Filippini in the Millar lab, is currently being used for displaying deep sequencing DNA and RNA data (Cell, May 2, 2008, http://neomorph.salk.edu/epigenome.html) and will be established for proteogenomic mapping of MS/MS spectral data in 2008/09.

- Meetings and the organization of proteomics workshops are established at a regular basis at the International Conference on Arabidopsis research (see ICAR http://www.plantconferences.org/Arabidopsis2008/). A workshop was held in 2007 in Beijing, and another is planned for Montreal in 2008.

- The establishment of large protein interaction databases and their cross-linking to MASC proteomics resources is the main goal of the next years.