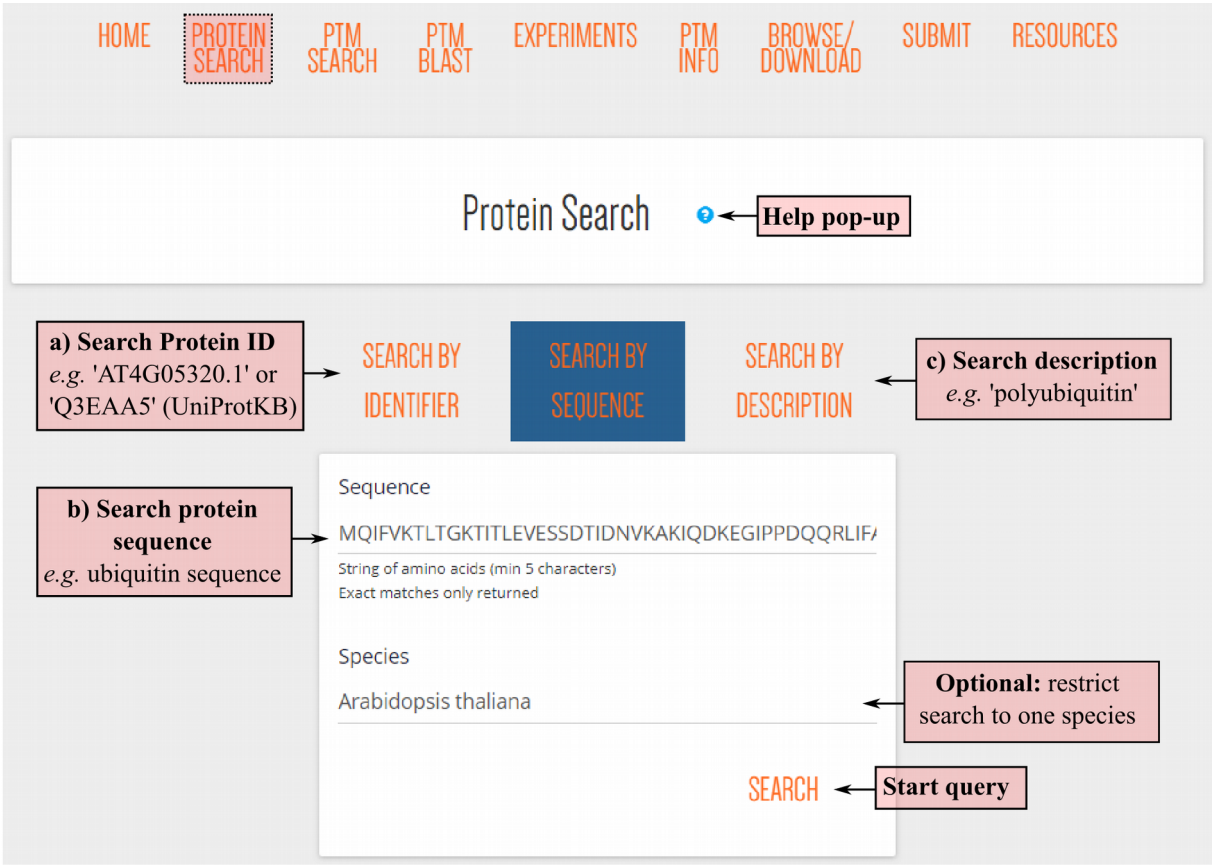
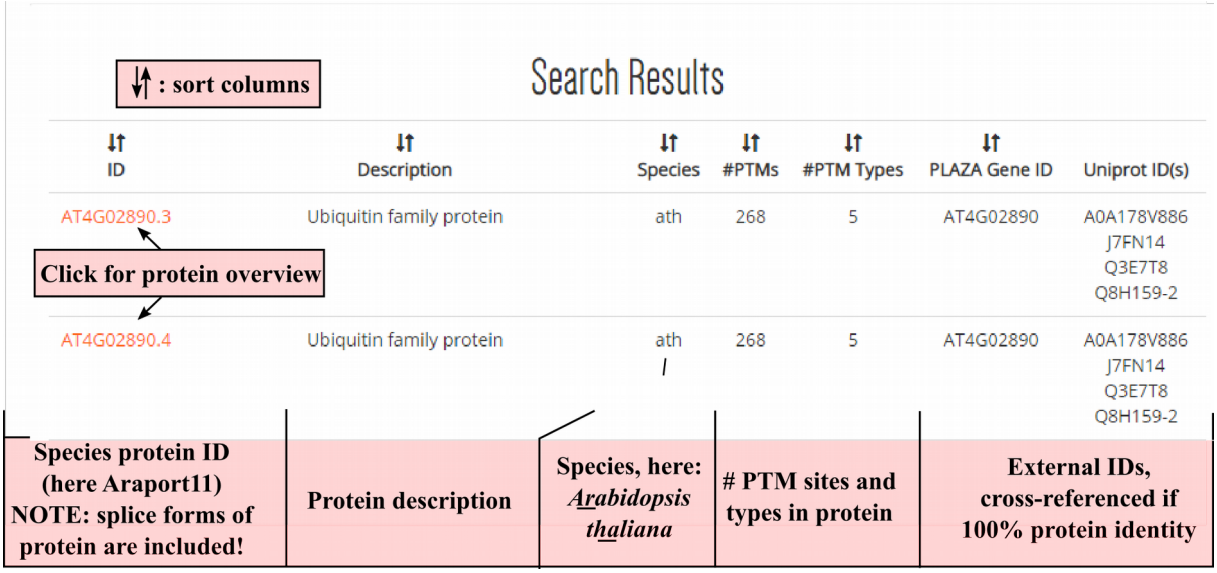
PROTEIN SEARCH TUTORIAL

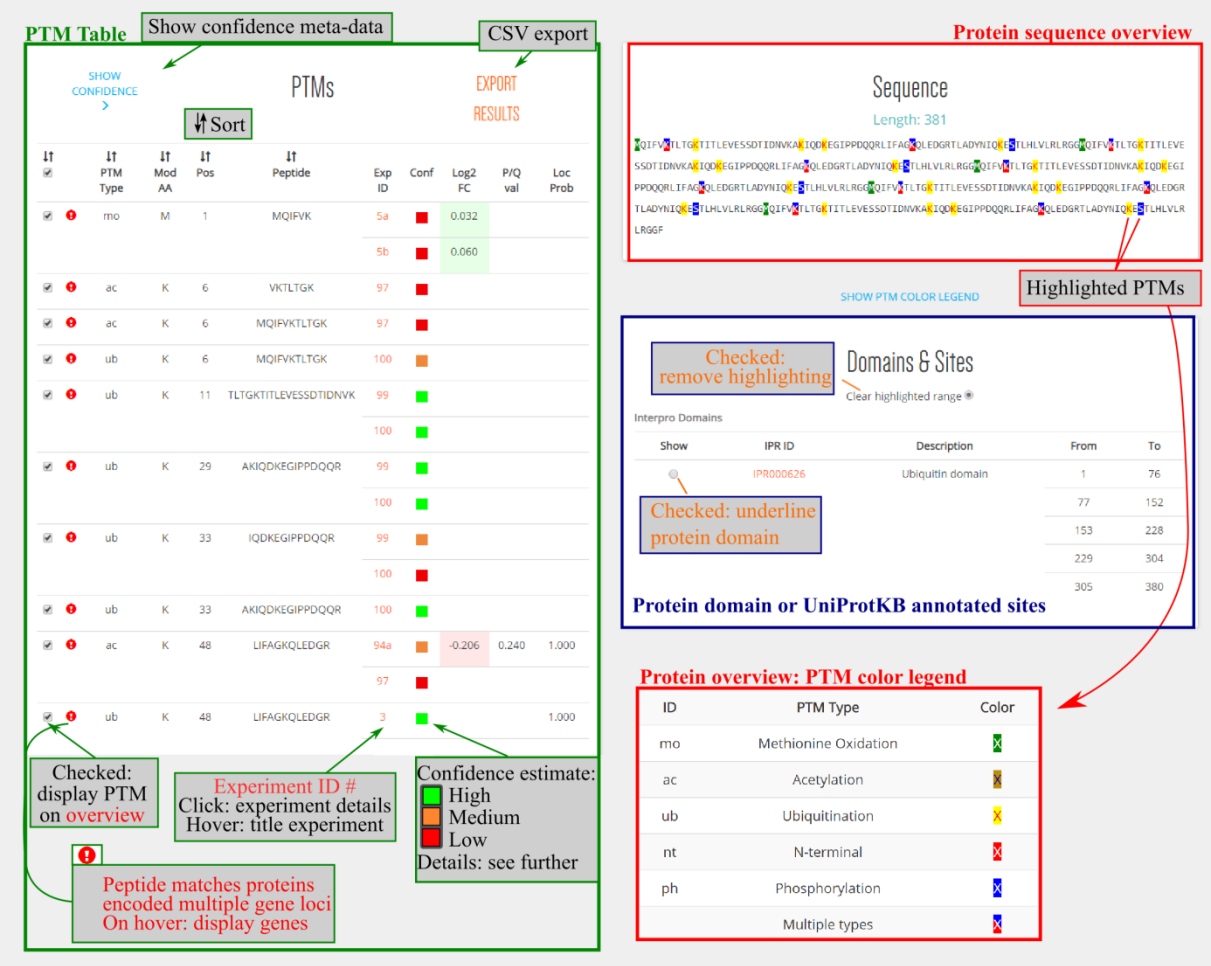
Users have three options to find a protein of interest: (a) search by protein identifier, (b) search by sequence (here ubiquitin), or (c) search by protein description. In addition, the search can be restricted to a single plant species by using the dropdown menu selection option below.



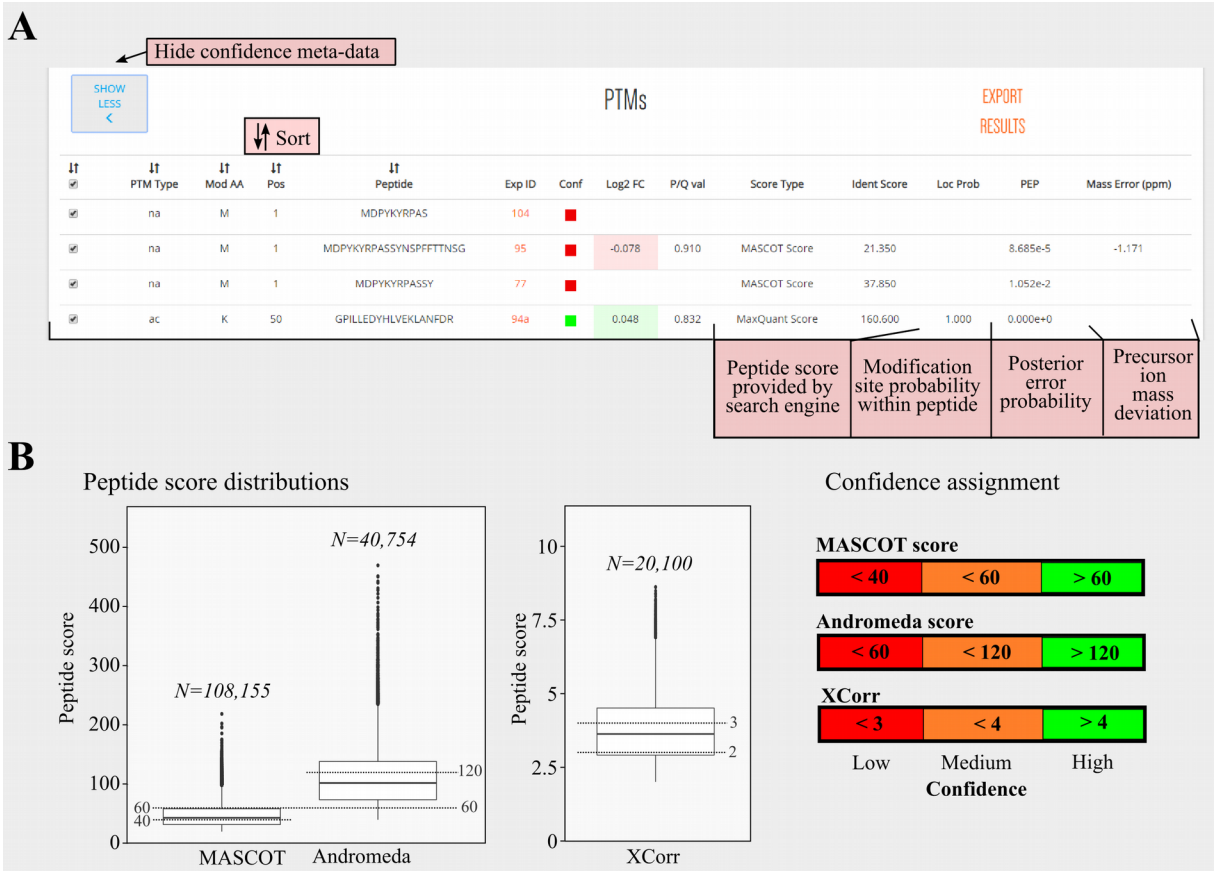
After pressing the ‘search’ button, any results will appear below the query box. All proteins fulfilling the criteria will be listed in the search results table. Note that the result table also includes protein splice forms! All columns can be sorted, including a description column, species abbreviation, cross-references protein identifiers of PLAZA or UniProtKB (requiring identical protein sequence) or the amount of PTM sites and types.



By clicking a protein identifier the PTM protein sequence overview is launched, as example we show here the protein encoded by polyubiquitin 10 (AT4G05320.1). Below a general protein info header with description and cross-references, a PTM table (left, green border), PTM protein sequence overview (top-right, red border) and protein domain/site table is provided (bottom-right, blue border). These are interactively connected to each other. For instance, by default all PTM checkboxes are selected in the PTM table. Removing a specific checkbox will remove the highlighting in the protein sequence overview. Note that a color legend can be displayed and also by hovering over a modified amino acid, the modification(s) will appear in a pop-up box. Similarly, a protein domain can be selected, e.g. here all ubiquitin domains were selected, and the domain will be underlined in the PTM protein sequence overview. In the PTM table additional information is found such as the type of PTM with corresponding protein position, the originating (plain) peptide identified by MS, the respective publication and a confidence color-coding. By clicking the MS study, the experiment overview is launched. If localization probabilities or differential abundance estimates (log2 fold change and significance) are available, these are displayed as well. Log2 fold changes are displayed in a heatmap-like gradient (green is upregulated, red is downregulated). In case the significance estimate was below the threshold employed in the respective study, this is also highlighted in green (note this was not the case here). The PTM table can be exported by clicking the 'Export results' button.



Details of the confidence meta-data collected can be consulted by clicking 'SHOW CONFIDENCE'. Below, we can view the extended version (figure panel A) including these confidence estimates reported by experiments, including peptide scores, posterior error probability (PEP), modification site localization probability and/or precursor mass deviation. Peptide scores are measured by search engines and score how a tandem mass spectrum matches a peptide from the searched protein database. Most frequently reported scores (used search algorithms) are the MASCOT ion score (MASCOT, Perkins et al., 1999), the Andromeda score (built-in MaxQuant software suite, Cox et al. 2011) and the cross-correlation score (XCorr, originally for SEQUEST, Eng et al., 1994). Distributions of these scores can be consulted in the figure panel B below. For these three search engines minimal peptide score thresholds were used. MASCOT ion scores were required to be at least 20, Andromeda scores 40 and XCorr scores at least 2. Next to peptide scores, which are highly differing and dependent on the search algorithm used, the PEP provides a more unified confidence estimate and can be considered as a "local FDR" that expresses the chance that a given peptide-to-spectrum match was incorrect. Most PEP values reported here were measured by software such as MaxQuant (Cox and Mann 2008), Proteome Discoverer (Thermo Scientific) or post-processing algorithms such as Percolator (Käll et al., 2007). Lastly, beside peptide-level confidence measurements, modification localization probability within a peptide can be assessed by algorithms such as PhosphoRS (Taus et al., 2011) or the PTM Score implemented in MaxQuant (Olsen et al., 2006). Here, we required a modification site localization probability of at least 0.75, when reported. Based on the peptide scores provided, PTMs are categorized as being low, medium or high confident (figure panel B - right). Assessing reliability of PTMs is a crucial step as false positive identification may occur in mass spectrometry identification results. In this aspect, careful inspection of experimental details remains therefor advisable.



REFERENCES

Cox, J., Neuhauser N., Michalski A., Scheltema R.A., Olsen J.V. and Mann M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794-1805.

Eng, J.K., McCormack, A.L. and Yates J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989.

Käll, L., Canterbury J.D., Weston J., Noble W.S. and MacCoss M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods **4**, 923-925.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648.

Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567.

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C. and Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **10**, 5354-5362.